

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 00:05:35 ; Search time 24.93 Seconds
(without alignments)
738.744 Million cell updates/sec

Title: US-09-837-602-2
Perfect score: 3899
Sequence: 1 MWKLLPAGGAGGEPYRLLT.....KEESLADDFRYNPLYKRRR 754

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	170.5	4.4	1939	US-09-310-187A-1	Sequence 1, Appli
2	169	4.3	1128	US-08-923-992A-6	Sequence 6, Appli
3	165	4.2	1886	US-08-938-105-3	Sequence 3, Appli
4	161	4.1	1098	US-08-923-992A-8	Sequence 8, Appli
5	158.5	4.1	1664	US-08-642-846-2	Sequence 2, Appli
6	158.5	4.1	1664	US-09-264-604-2	Sequence 2, Appli
7	157.5	4.0	976	US-09-104-324B-4	Sequence 4, Appli
8	157	4.0	1164	US-08-923-992A-2	Sequence 2, Appli
9	156	4.0	1164	US-08-923-992A-10	Sequence 10, Appli
10	152.5	3.9	754	US-09-214-564A-2	Sequence 2, Appli
11	150	3.8	461	US-08-630-822A-68	Sequence 68, Appli
12	150	3.8	461	US-09-005-069-68	Sequence 68, Appli
13	146.5	3.8	1093	PCT-US93-03077-1	Sequence 1, Appli
14	146.5	3.8	3248	US-08-353-700-1	Sequence 1, Appli
15	146.5	3.8	3248	PCT-US95-16216-1	Sequence 1, Appli
16	145	3.7	1702	US-08-296-791-5	Sequence 5, Appli
17	145	3.7	1702	PCT-US95-10661A-5	Sequence 5, Appli
18	144.5	3.7	1014	US-09-078-347A-3	Sequence 3, Appli
19	144	3.7	1588	PCT-US93-07261-11	Sequence 11, Appli
20	144	3.7	1663	PCT-US93-07261-16	Sequence 16, Appli
21	143	3.7	904	US-09-198-484-2	Sequence 2, Appli
22	143	3.7	1013	US-08-860-886-2	Sequence 2, Appli
23	142.5	3.7	2482	US-08-328-254-6	Sequence 6, Appli
24	141.5	3.6	1120	US-09-147-404-1	Sequence 1, Appli
25	141.5	3.6	1354	US-08-685-871-2	Sequence 2, Appli
26	141	3.6	1494	US-08-755-587-186	Sequence 186, App
27	140	3.6	1435	US-08-568-459A-4	Sequence 4, Appli

28	140	3.6	1435	2	US-08-487-826B-4	Sequence 4, Appli
29	139.5	3.6	984	1	US-08-242-932-2	Sequence 2, Appli
30	139.5	3.6	984	1	US-08-714-481-2	Sequence 2, Appli
31	139.5	3.6	984	5	PCT-US95-06111-2	Sequence 2, Appli
32	139	3.6	1104	4	US-08-923-992A-4	Sequence 4, Appli
33	138.5	3.6	1786	4	US-08-973-462-8	Sequence 8, Appli
34	136.5	3.5	3788	4	US-09-336-447A-76	Sequence 76, Appli
35	136	3.5	1075	2	US-08-993-228-19	Sequence 19, Appli
36	135.5	3.5	683	6	5210183-3	Patent No. 5210183
37	135.5	3.5	1201	4	US-09-098-901-2	Sequence 2, Appli
38	135	3.5	1036	4	US-08-891-640-3	Sequence 3, Appli
39	134	3.4	1184	4	US-09-541-782-2	Sequence 2, Appli
40	134	3.4	2285	4	US-09-308-375-2	Sequence 2, Appli
41	133	3.4	376	6	5180810-1	Patent No. 5180810
42	131.5	3.4	558	1	US-08-285-440-6	Sequence 6, Appli
43	131.5	3.4	558	1	US-08-630-349-6	Sequence 6, Appli
44	131.5	3.4	2343	4	US-09-324-867-2	Sequence 2, Appli
45	131	3.4	595	1	US-08-171-718-16	Sequence 16, Appli

ALIGNMENTS

RESULT 1
US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751
; GENERAL INFORMATION:
; APPLICANT: Benichou, Gilles
; APPLICANT: Fedoseyeva, Eugenia
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
; FILE REFERENCE: UCSF-090
; CURRENT APPLICATION NUMBER: US/09/310,187A
; CURRENT FILING DATE: 1999-05-12
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-310-187A-1

Query Match	4.4%	Score 170.5;	DB 4;	Length 1939;
Best Local Similarity	23.4%	Pred. No. 8 6e-05;		
Matches	91;	Conservative	57;	Mismatches 132; Indels 109; Gaps 17;
QY	388	KMQKFRMLSQDAPTVKESCKTSSNNNSMVSNTLAKMRIPNYOLSPKLPKSKDRAS	447	
DB	833	KLYFKIKPLKLSAETKEMA-TMKEEFGRKTKETLEKSEARRKELE-EKMYSLQKENDLQ	890	
QY	448	QQOOTS-----IRNYFQPTKKR---ERDEENQEMS---SCKSARIETSCS	488	
DB	891	LQVQAEQDNLDAERCDQLIKNKIOLEAKVKEMNERLEDEEEMNAELTAKRKLEDCS	950	
QY	489	-LLEQTOPATPPLGWK-NKEOHLSENPEVDNTSDNNLFTDRLSKSVKNSKSHAAEKLK	546	
DB	951	ELAKDIDDLTLAKVEKEKHATENK-----VKNLTSEMAGLDEIIAKLT	995	
QY	547	SNKK-----REMDVDAIDEVLEQLFKDTPKELEIDVKVQOEEDVNYKRRPRMDIET	599	
DB	996	KEKKALQEAHQALDDLQVEDKVNLSL---SKSKVLEQQVDDLEGSLEQEKVKVMDLER	1052	
QY	600	NTFSDAENVPESSKISQENIGKREKLEKEDSLWSAKEISN--NDKLODDSEMLPKKLLLT	657	
DB	1053	-----AKRLEGDLKLTQESIMDLNDKLO-----LEEKLKKK	1085	
QY	658	EFKSLYIKNSTRNPSGINDDYG-OLKNFKKKFKVTPYGAGKPLHIITGGSLTAHARKN	716	
DB	1086	EF-----DINQONSKIEDEQALQLOKRLKE-----NOARI	1117	
QY	717	TELEWLRQEMEVQNOHAKESLADDLFR	745	

Qy	192	QIESFVPPDPEPSIGKNVDLSGR-----QERKOLFCKTTFINAKOHHKLSSA	241
Db	95	QKNEFKTKIDETDNDSDALLELQNFNETNRLLLHIKOEVEEKD-----KAKQOKTLKOS	149
Qy	242	VFEGGEARLITEENEEHNFFLAPGTCVVD-TGIITSQTLIPDCOKKWTOSTMDMLQK	300
Db	150	-----DTKVDLSNIDKELNHQKSPVEKNAEPKGIINED---KDSMLKIEDIRKQAQQA	200
Qy	301	GLRIPAEAI--GLAVIFMTTKYCDQHPSTGLKTTTTPGSPSOGVSVDEKMLPSAPV	358
Db	201	DKKEDAEVKVREELGLFSSTKAGLQOEIHEHVKKETS---SEENTQKVDEHY-----A	251
Qy	359	NTTYYVADTSEQADTWDLSEPKETIKVSKME--QK-----FRLMSQDAP	401
Db	252	NSLQNLAKSLEELDKNATTNEQATQVKNQFLENQAOKLKEMQPLIKETNVKLYKAMESLE	311
Qy	402	TVRESCKTSSNNN--SWGSNTLAKMRIPNYQLSPTK-LPSINKSKDRASO--QQQNTSIR	456
Db	312	QVEKLKHNSEANLEDAVAKSEIIVREYEGKLNQSKNLPKELKOLEEAHSKILQVYVEDFR	371

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match 4.2%; Score 165; DB 4; Length 1886;
Best Local Similarity 22.5%; Pred. No. 0.00023;
Matches 91; Conservative 61; Mismatches 123; Indels 13

QY 598 ETNDTSDEAVPSSKISQENEIGKREKEDSLWSAKEISNNDKLQDDSEMLPKKLLLT 657
Db 571 ALNDDESDTQNTKMSRFHI-----DSDWLEDSNGDRDND-----IS 614
QY 658 EPRSIVKINSTSRNPSSINDDYQOLNFKKKVTPPGACKLPHIIGGSDLIHAHAKNT 717
Db 615 RFKSDILNDVSTSDIIGDKYNSSEITTKTLAPRS-----DNNDKENSKSLEDP 667
QY 718 ELEEWLRQEMEVONQHAKEE 737
Db 668 ANNESLQQOLEV--PHTKED 685

RESULT 7
US-09-104-324B-4
; Sequence 4, Application US/09104324B
; Patent No. 6232460
; GENERAL INFORMATION:
; APPLICANT: T reci, Ozlem; Sahin, Ugur; Pfreundschuh, Michael
; TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,
; TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample Of
; TITLE OF INVENTION: No. 6232460mal Cells
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski LLP
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10103

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,324B
; FILING DATE: 25-June-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/892,702
; FILING DATE: 15-July-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6232460man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5491
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 318-3000
; TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 976 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-09-104-324B-4

Query Match 4.0%; Score 157.5; DB 4; Length 976;
Best Local Similarity 19.0%; Pred. No. 0.00035;
Matches 159; Conservative 150; Mismatches 312; Indels 215; Gaps 41;

QY 22 VEYVGRKNCALIIENDQSIQRHVAULTANFSVTNLQSDTEIPVLTKDONS-----X 73
Db 46 LEPPFAKTN---LSKNGENIDSPALQKVN-----LPVLEQVGNSDCHYQEGLK 92
QY 74 YGTFVNEEKMQNGFSRTLSKSGDITGVFGSKPRIEYE-PLVACSSCLDVSGK--TALNQ 130
Db 93 DSDLENSGLSRVFSKLYKEAEKI-----KKWVSTEAELRQKESKQLENRRKIIIEAQRK 146
QY 131 AI--LQLGGFTVNNWTEE--CTHLVMSVVKVTIKTICALI---CGRPVVKP-----EYFTEF 180
Db 147 AIQELQFGNEKVSLEEGIQENKDLIKENNATRHLCNLLKETCARSAEKTKKYVEREE 206
QY 181 LKAVQSKQPPQIESPYPLDEPSIGSKN-----VDLSGRQERKQIFKGTFFIL 230

Db 207 TRQVYMDLN--NTEKMITAFQELRVQAEHSRLEMHFKEDYEKIQHLEQEKKE-----I 261
QY 231 NAKQHKLLSSAVVGGGEARLITEENEEHFFLAPGTCVVDGTCITNSQTLIPDCQKKWI 290
Db 262 NDKK-KQVSLLLI-----QITEKENMKMDLTF-----LEESRKVNQOLEK-- 302
QY 291 QSIMDMLQROGLRPIPEAIEGLAVIFMTTKNYKCDPOGHPSTGLKTTTPGSLSGQSVDE 350
Db 303 -----TKLQSENL-----KQSTIEKHHLTKLELDIK--VSLQRSVSTQK 339
QY 351 KLMPSPAPVNTTYVADTSEQ-----ADTWLSE-----RPKEIKVS 387
Db 340 ALBEDLIQATKTKTCQTEKEKETQMEESNKARAAHSFVVTEFTTVCSLEELLRTQOORLE 399
QY 388 KMEQKFRMLSOD----APT VKESCKTSSNNN---SMVSNFTLAKMRIPNY-----Q 430
Db 400 KNEQDKILTMELQKKSSSELEEMTKLTNNKEVELEELKVLGKEKITLLYENKQFEKTAEE 459
QY 431 LSPTKLPSIN-----KSKDRASQOQNTSI---RNYFQPTKKRDERDENQEM-----SS 477
Db 460 LKQTEQELIGLLQAREKEVHDLEITQLTAITTSEQYYSKEVKDLKTELENEKLNKNTLTS 519
QY 478 CKSARIETSCSLEQTPATPSLWKNKEQHLSENEPVDNNS-----DNNLFTDLDLSIVK 533
Db 520 CNKLSLENK-ELTQETSDMTLEL-KNOQEDINNKKOEERMLKQIENLQETETQLRNELE 577
QY 534 -----NSAKSHAAEKLRSNKKREMDDDVAIDEVLEQLFKDKTPELE 575
Db 578 YVREELKQKRDVCKLDKSEENCNNLRKOVENKNKYIEELQENKALKKGTAEKQNLN 637
QY 576 I-DVKYQKQEDYVNRKRPRMDIETNDFSDAEVPPESKISQEN---EIGKKRELKBDLS 631
Db 638 VYEIKVKNLELESQAKQFGEI--TDYQKEI--EDKKISEENLLEVEKAKVIADEAV 693
QY 632 WSAKEISN--NDKLQDDSEMLPK-----KLLITEFRSLVKKTSRNPSSINDDYG-QL 682
Db 694 KLOKEIDKRCQHKIAEMVALMERKHQYDKIIEERDSELGLYKSKEOQSLSRASLEIEL 753
QY 683 KNEFKFKKVTPGAGKPLPHIIGGSDLIHAHARKNTELEEMLRQEMEVONQHAKEES 738
Db 754 SNLK-----AELLS--VKQLEIE---REEKELKREAKENT 785

RESULT 8
US-08-923-992A-2
; Sequence 2, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-Iga Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996

ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1164 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
us-08-923-992A-2

Query Match 4.08; Score 157; DB 4; Length 1164;
Best Local Similarity 20.18; Pred. No. 0.00049;
Matches 144; Conservative 121; Mismatches 236; Indels 216; Gaps 40;
Qy 192 QIESFYPLDPSIGSKNVDLSGR-----QERKQIFKGTFFLNAKQHKLSSA 241
Db 131 QKNEFKTIDETNSDALLELENFNETNRLHLIKQHEVEKDK-----KAKQKTLKQS 185
Qy 242 ---VFGGGEARLITEENEEHNFPLAPGTCVDTGINTSOTLIPDCOKKWIQSIMDLQ 298
Db 186 DTKVDLSNIDKELNHQKSOVE-----KMAEQKGITNED---KDSMLKKIEDIRKQAO 234
Qy 299 ROGRLRPIEAET--GLAVIEMTTKNYCDP--QGHSTGCTKTTTPGPSISOGVSVDEKLMP 354
Db 235 QADKEDAEVREELKLFSTAGLDQEOEH-----VKETSEENTQ--KVDEHY-- 286
Qy 355 SAPVNTTYVADTSEQADTWLSERPKEIKVSKME--QK-----FRMLS 397
Db 287 ---ANSQNLAKQSLEELDATTNEQATQVKNOFLENAQKLEIQPLIKETNVKLYKAMS 343
Qy 398 QDATVRESCKTSSNNN--SMVSNLTAKMRIPNLTOLSPK--LPSINKSKDRASQ--QQOT 452
Db 344 ESLEQVEKELHNEANLEDLVAKSKIEVREYEGKLNQSKNLPKQLQEEAHSKLKQOV 403
Qy 453 NSIRNYF----QPSTKKR-ERD-----BENOEMSSCKSARIETSCSLLEQTOPATPSSLWN 503
Db 404 EDFRKKFKTSEQVTPKRVKRDLAANENQO-----KIELTVS-----PENITVYEG 450
Qy 504 KEQHLSENEPVDTSNNLFTD--TDLKSIYKNSASKSHAABKLRNKKREMDVAIEDEV 562
Db 451 EDVAKFT----VTAKSDSKTTLDFSDL--LTKYNPSVS---DRISTNYKTNTDNHKAIEIT 501
Qy 563 LEQL-----FKDTK---PELEIDVKVQKEEDVNVKRRPRMDIETNDFESDEAV 608
Db 502 IKNLKLNESQVTLKAKDSDGNVVEKFTTITVQKKEEK--QVPKTP-----EQKDSKTEKV 556
Qy 609 PESSKISOENET-----GKKRELK-----EDSLW-SAKE--- 636
Db 557 PQEPKSNKQNLQELIKSAQOELEKLEKAIKELMEQPEIPSNPEYGIQKSIWESQKEPIQ 616
Qy 637 --ISNNDKLQDSS-----EMLPKLLL-----TEFR 660
Db 617 EAITSFKKIIGDSSSKYTEHYFNKYSKDFMNYQLHAQMEMLTRKVQVYMNKYPDNEAIK 676
Qy 661 SLVIKNSTSRNPISNDYDGLKN-----FKKP-----KKVTPYAGK 698
Db 677 K-IFESDMKRTK----EDNYGLENDALKGFEKYLTPFNKIKQIVDDLKKKVEDQDQAP 732
Qy 699 LPHIIGSDLTAHARKNTELEEWLROEMEVONQH--AKEESLADDLFRYPNPKRR 753
Db 733 IPE---NSEMDQAEKAKIAVSKVSKVLDGVHQLKKNKSNKIVDLFKELEAIKQO 786

RESULT 9
us-08-923-992A-10
Sequence 10, Application US/08923992A
Patent No. 6280738

GENERAL INFORMATION:
APPLICANT: Tai, Joseph Y.
APPLICANT: Blake, Milan S.
TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
TITLE OF INVENTION: Streptococcal Beta Antigens
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,992A
FILING DATE: 05-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,707
FILING DATE: 06-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1164 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
us-08-923-992A-10

Query Match 4.08; Score 156; DB 4; Length 1164;
Best Local Similarity 19.48; Pred. No. 0.00059;
Matches 133; Conservative 125; Mismatches 239; Indels 212; Gaps 37;
Qy 192 QIESFYPLDPSIGSKNVDLSGR-----QERKQIFKGTFFLNAKQHKLSSA 241
Db 131 QKNEFKTIDETNSDALLELENFNETNRLHLIKQHEVEKDK-----KAKQKTLKQS 185
Qy 242 ---VFGGGEARLITEENEEHNFPLAPGTCVDTGINTSOTLIPDCOKKWIQSIMDLQ 298
Db 186 DTKVDLSNIDKELNHQKSOVE-----TMAEQLGITNED---KDSMLKKIEDIRKQAO 234
Qy 299 ROGRLRPIEAET--GLAVIEMTTKNYCDPQGHSTGCTKTTTPGPSISOGVSVDEKLMP 356
Db 235 QADKEDAEVREELKLFSTAGLDQEOEHVKRETT-----SEENTQKVDEHY----- 286
Qy 357 PVNTTYVADTSEQADTWLSERPKEIKVSKME--QK-----FRMLSQD 399
Db 287 -PNSQLNLAOKSLEELDATTNEQATQVKNOFLENAQKLEIQPLIKETNVKLYKAMS 345
Qy 400 APTVRESCKTSSNNN--SMVSNLTAKMRIPNLTOLSPK--LPSINKSKDRASQ--QQOTNS 454
Db 346 LEQVEKQLKHNSQANLEDLVAKSKIEVREYEGKLNQSKNLPKQLQEEAHSKLKQOV 405
Qy 455 IRNYF---QPSTKKR-ERD-----BENOEMSSCKSARIETSCSLLEQTOPATPSSLWN 505
Db 406 FRKKFKTSEQVTPKRVKRDLAANENQO-----KIELTVS-----PENITVYEG 452
Qy 506 QHLSENEPVDTSNNLFTD--TDLKSIYKNSASKSHAABKLRNKKREMDVAIEDEV 564
Db 453 LKFT---LTAKSDSKTTLDFSDL--LTKYNPSVS---DRISTNYKTNTDNHKAIEIT 503

Matches 93; Conservative 70; Mismatches 155; Indels 98; Gaps 21;

QY 270 VVDFGITSOTLIPD-CQKKWIOSIMDLQROGLR-----PIPEAEIGLAVIFMTTKN 321
Db 43 MLDPLNPNQOLMFNYMQLQLOLQELHLSQQQPMHFEHPIQOE-----ATSTN 94

QY 322 YCDPQHPSGLKTTTPGSPSLSQGVSVDEKMLPSAPV-NTTTYVADTSE-----QA 372
Db 95 Y-GPSGOYITSDATSYO-----SIAQOFVPOPPITETTTKIPETEIQIGVSNQYQON 145

QY 373 DTWDLSPRKEIKVSKMEQKFR-----MLSODAPTVKESCKTSSNNNSMVSNNT- 420
Db 146 ITNSNISPEVI-----GFREHYVAEQSGDVLHSHLHTEQPADKSTRGDQBPVSETG 198

QY 421 --LAKMRIPNYQLSPTK-LPSINKSKDRASQQOQTNSIRNYFQSTKKRDERDEENQEMSS 477
Db 199 SGFSYAQILSQGLKPTQPSNSVNLADRSPDLDTKTKENYKSP-----242

QY 478 CKSARIETCSLLEQTOPATPSL-WKNKEQHLSENEPVDNNSDNLFTDLDKLSIVKNSA 536
Db 243 ---GRVQDITKIIDEKOKSKDTEWHNKK--VKEHKKVKD-----IKPDFESSQRNKK 290

QY 537 SKSHAAEKLRSNKKREMDVAIEDEVL--EOLFKDTPKELEIDVKVQOEEDVNVKRPR 594
Db 291 SKN--IPKQIENTPQLDSLRSDIVIKGELLTKDTTKSL--TTVNVVSELDVSKPKDEKP 347

QY 595 MDIETNDTFSDEAVPES-SKISQENEIGKKRELK-----EDLSWSA-KEISNNDK 642
Db 348 EPSEPSKTFIDTSVAKVDNDSTQANHKKKSKSPKRPTEPEDEIEKALKETQASEK 403

RESULT 12

US-09-005-069-68
; Sequence 68, Application US/09005069
; Patent No. 5932470
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU
; APPLICANT: WALLENFELS, LYNDIA
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
; TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/630,822
; FILING DATE: 11-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: CONNELL, GARY J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-17-C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 461 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

; MOLECULE TYPE: protein
US-09-005-069-68

Query Match 3.8%; Score 150; DB 2; Length 461;
Best Local Similarity 22.4%; Pred. No. 0.00046;
Matches 93; Conservative 70; Mismatches 155; Indels 98; Gaps 21;

QY 270 VVDFGITSOTLIPD-CQKKWIOSIMDLQROGLR-----PIPEAEIGLAVIFMTTKN 321
Db 43 MLDPLNPNQOLMFNYMQLQLOLQELHLSQQQPMHFEHPIQOE-----ATSTN 94

QY 322 YCDPQHPSGLKTTTPGSPSLSQGVSVDEKMLPSAPV-NTTTYVADTSE-----QA 372
Db 95 Y-GPSGOYITSDATSYO-----SIAQOFVPOPPITETTTKIPETEIQIGVSNQYQON 145

QY 373 DTWDLSPRKEIKVSKMEQKFR-----MLSODAPTVKESCKTSSNNNSMVSNNT- 420
Db 146 ITNSNISPEVI-----GFREHYVAEQSGDVLHSHLHTEQPADKSTRGDQBPVSETG 198

QY 421 --LAKMRIPNYQLSPTK-LPSINKSKDRASQQOQTNSIRNYFQSTKKRDERDEENQEMSS 477
Db 199 SGFSYAQILSQGLKPTQPSNSVNLADRSPDLDTKTKENYKSP-----242

QY 478 CKSARIETCSLLEQTOPATPSL-WKNKEQHLSENEPVDNNSDNLFTDLDKLSIVKNSA 536
Db 243 ---GRVQDITKIIDEKOKSKDTEWHNKK--VKEHKKVKD-----IKPDFESSQRNKK 290

QY 537 SKSHAAEKLRSNKKREMDVAIEDEVL--EOLFKDTPKELEIDVKVQOEEDVNVKRPR 594
Db 291 SKN--IPKQIENTPQLDSLRSDIVIKGELLTKDTTKSL--TTVNVVSELDVSKPKDEKP 347

QY 595 MDIETNDTFSDEAVPES-SKISQENEIGKKRELK-----EDLSWSA-KEISNNDK 642
Db 348 EPSEPSKTFIDTSVAKVDNDSTQANHKKKSKSPKRPTEPEDEIEKALKETQASEK 403

RESULT 13

PCT-US93-03077-1
; Sequence 1, Application PC/TUS9303077
; GENERAL INFORMATION:
; APPLICANT: Board of Regents, The University of Texas System
; APPLICANT: Gaynor, Richard B.
; APPLICANT: Wu, Foon Kin
; TITLE OF INVENTION: PROTEIN CELLULAR FACTOR USEFUL FOR
; TITLE OF INVENTION: REGULATING GENE EXPRESSION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03077
; FILING DATE: 19930331
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,025
; FILING DATE: April 2, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Kammerer, Patricia A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: UTFD270PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1540
; TELEFAX: 713-749-2679

; TELEX:

```

; INFORMATION FOR SEQ ID NO: 1:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 1093 amino acids
;
; TYPE: AMINO ACID
;
; STRANDEDNESS: unknown
;
; TOPOLOGY: unknown
;
; MOLECULE TYPE: protein
;
PCT-US93-03077-1

```

Query Match	3.8%;	Score 146.5;	DB 5;	Length 1093;
Best Local Similarity	17.5%;	Pred. No. 0.0031;		
Matches 142;	Conservative 135;	Mismatches 272;		
		Indels 261;		Gaps 35;

QY	47	VLTFANFSVTNL	SO---	TDEIPVLTLKONSKYGT	FPVNEEK	MONGRS-	---RTLKSGDITFG	100
Db	203	VKTMTESIS	INTSOTSLQ	TAETKIDALEKEQ	----	KHEDQ	SNTSPPVSTFSSG	258
QY	101	VFGSKPRIEY	EPLVACSSCLD	VSCKTALNQAILQ	LGGFTYNNW	TTECH	TLVWYSKVYTIK	160
Db	259	I----	EVLDES	VISESASSRQ	TTDSKSL	-----	HLMQTFQLLSA	298
QY	161	TICALICGR	PIVPEY	-----	FTEFLKAVOSK	KOPPQIESF	-YPLDPSIGSNVD	212
Db	299	SAC	-----	PEYNRLDD	FOKLTESCCSSDA	FERIDSFV	SQLSRSYSEINS	348
QY	213	SGRQERKQI	PKGTFIFL	NAKOHKKLSAV	VFGGGEARL	ITERNEE	HINFLEAPGTCVVD	272
Db	349	SG	-----	KGYALVPII	VNSSTPKTS	AEBSGKEG	EEVNE-----	386
QY	273	TGITNSOTL	IPDCQKKW	IQSIMDLQ	ROGLRPIE	AEIGLAVI	FWTTKNYCDPOGH	332
Db	387	-----	-----	-----	PTEAEM	-----	-----EESGRSATP	402
QY	333	LKTTTPGSL	SQGSVSVDEK	LMPSPAVNTT	VYADTESQ	ADTWLSE	RPKIKYSKMBQK	392
Db	403	VNCEQP	-----	DILVSTPI	NEGOTVLDK	VAEQE-----	PAESQPEALSEK	444
QY	393	FRMLSQDAP	TVKESCTSSNN	SNMVSNTLAK	MRIPNYQLSP	TKPLPSIN	KSKDRASOQOQT	452
Db	445	-----	EDCKT	-----	VEFLNEKLEK	REAGLJLS	KEKALLLEAF	479
QY	453	NSIRNYP	-----	QPSTKKRER	DEENQEMSSCK	SARIETCS	LLQOTOPATPSLW	507
Db	480	DNLKDEMFR	YKEESSISL	DKFTQRIAEAK	-KVQLACKER	DAAKKEIK	NIKEELATR	538
QY	508	LSENEPVT	NSDNNLFTD	TLKSIV	-----	KNSASKSHAAE	KLRSKKKREMDVA	557
Db	539	LNSSETADL	LKEK	-----	DEQIRGLME	ESEKLSKQOLH	NSNIKKLRADKDKEN	594
QY	558	-----	IEDEV	-LEOLF	KDTPKELE	-----	IDVKVOKOEDVNVVR	597
Db	595	KYKELEEE	HLQHLKQVL	-DKGEEV	EKOHRENI	KKLNSMVER	QEKDLG---	650
QY	598	ETNDTFS	-----	DEAVPE	-----	SSKISQBN	EICKRELK	645
Db	651	EKNRSIQAA	LDSAYKELT	DLHKANA	AKDSEQA	EAAALSR	EMKAKEELSAA	710
QY	646	DSEM	-----	-----	-----	LPKLLLT	FERSLVINKSTS	669
Db	711	QETLAIQV	DLRLALQ	RTEQA	AAKEDYLR	HEIGELQ	ORLOEARNNOELS	770
QY	670	--RNPSG	INDDYG	-OLKNFK	FKFKVTVP	GACKLPHI	IGGSD--	724
Db	771	LLRQIENL	QATLGSQ	TSSWEKLEK	-----	NLSDR	LGESQTLAA	823
QY	725	QEMEVO	QNOHAK	EESLAD	DLFRYNP	YLKRRR	754	
Db	824	NKIOM	SWESQNSL	LRQENS	RFQALRSEK	853		

RESULT 14

```

US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DOREFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

```

Query Match 3.8%; Score 146.5; DB 1; Length 3248;
Best Local Similarity 18.7%; Pred. NO. 0.015;
Matches 152; Conservative 136; Mismatches 314; Indels 211; Gaps 33;

Qy	34	LIENDQISIRNHAVLFANFSVNLNLSOTDEIPVLTLKDNKSYGTFVNE-EKMQNGFSTLK	92
Db	2129	LQEKILQSLKDSQAL----SLTKCELENOIAQL----NKEKELLAVKESIESIQAARLSBS--	2178
Qy	93	SGDGIIFGVGSKFRIEYEPYLVACSSCLDVSKTALNQAIIQLGGFTVN-NWTEECTHLV	151
Db	2179	-----DYEK-----LNVS-----KALEAALVEKGEFALRSLSTQOEYHQ	2213
Qy	152	--MVSVKVTIKTICALICGRPIVKPEYFTEFLKAVQSKQ--PPOIESFYPPIDEPSIG	206
Db	2214	RRGIEKURVIEA-----DEKKOLHIAEKLKERENDSLDKDVENLEHRELOMSEEN	2265
Qy	207	SKNVDLSGROERQIFKGKTFIFLNAQKHKLSSAVFGGGEARLITENEBEENHFFLAP	266
Db	2266	QELVILDAENSKAEVETLKTQIEEMARSLKIFELDLYLRSEKNLTQIOEKQG-----	2320
Qy	267	GTCVVDTGINTSLTIPDCQKKWQ-----SIMDMLQRGLRPIPEAEIGLA---VIFM	317
Db	2321	QLSELDKLLSFRKSLLEEKAEQAEIQIKESKTAIVEMLQNO-LKELNEAVALCGDQEI	2379
Qy	318	TTKNYCDPOGHPSTGLKTTTPGDSISQGVSVDBKL-----	352
Db	2380	ATEOSIDPPTIEEBHOLRNST--EKURARLEADBEKKOLCVLOOLAKESHHADLLKGRVENL	2437

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:26:30 ; Search time 3146.82 Seconds
(without alignments)
18884.809 Million cell updates/sec

Title: US-09-837-602-1

Perfect score: 4403

Sequence: 1 ttccgacagcgcggttg.....accggtggagctccagct 4403

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthba:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1460.2	33.2	1497	11 BC005293	BC005293 Homo sapi
2	1427	32.4	1460	11 BC016762	BC016762 Homo sapi
3	1391	31.6	2518	11 BC003719	BC003719 Mus muscu
4	1090.4	24.8	1792	11 BC013190	BC013190 Mus muscu
5	822.8	18.7	927	10 BG182890	BG182890 RST1768 A
6	741	16.8	789	10 BG214621	BG214621 RST34264
7	732.4	16.6	902	10 BM461758	BM461758 AGENCOURT
8	725.4	16.5	775	10 BG676775	BG676775
9	706.6	16.0	763	10 BM014420	BM014420 603640169
10	696.2	15.8	960	10 BG284646	BG284646 602408824
11	690.8	15.7	779	10 BG388866	BG388866 602414591
12	687.8	15.6	712	9 AI796269	AI796269 wh44g10.x
13	676.2	15.4	735	10 BG779728	BG779728 602668556
14	672.6	15.3	807	10 BG197194	BG197194 RST16433
15	671	15.2	752	9 AU118357	AU118357 AU118357
16	669.2	15.1	718	9 AW976050	AW976050 EST388159
17	663	15.1	845	10 BG532783	BG532783 602580405

C	18	660.6	15.0	691	9	AW978306	AW978306 EST390415
	19	651.2	14.8	706	10	BG292394	BG292394 602386511
	20	648.4	14.7	690	10	BG506358	BG506358 601860423
	21	644	14.6	928	10	BG392111	BG392111 602410095
	22	633.4	14.4	686	10	BG483074	BG483074 602502895
	23	629	14.3	637	9	AW237021	AW237021 xm52b10.x
	24	627.2	14.2	881	10	BG109073	BG109073 602281453
	25	621.2	14.1	667	10	BF222791	BF222791 7q23a10.x
	26	612.6	13.9	642	10	BG613414	BG613414 602641314
	27	608.6	13.8	642	9	AI888159	AI888159 wm52h04.x
	28	600.8	13.6	889	10	BE786964	BE786964 601476452
	29	596	13.5	604	9	AW183153	AW183153 x167b11.x
	30	589	13.4	909	10	BF027776	BF027776 601764083
	31	587.4	13.3	1024	10	BG202556	BG202556 RST21915
	32	581.4	13.2	605	10	BF511289	BF511289 UI-H-B14-
	33	576.4	13.1	672	10	BG527527	BG527527 602557140
	34	571.6	13.0	575	10	BM312264	BM312264 1q41c07.y
	35	569	12.9	728	10	BF382172	BF382172 601814949
	36	564.2	12.8	651	9	AV716658	AV716658 AV716658
	37	559.2	12.7	872	10	BF573240	BF573240 602079623
	38	558.6	12.7	593	10	BF219347	BF219347 601884125
	39	554.8	12.6	573	10	BE694454	BE694454 QV0-BT070
	40	553.4	12.6	555	10	BI962748	BI962748 1e54b06.y
	41	553.4	12.6	789	10	BI691051	BI691051 603309822
	42	543.8	12.4	636	9	AW363125	AW363125 CM2-CT031
	43	537.8	12.2	708	10	BG530983	BG530983 602560854
	44	537.6	12.2	757	10	BF028917	BF028917 601765293
	45	537.4	12.2	796	10	BF248479	BF248479 601821553

ALIGNMENTS

RESULT	1	1497 bp	mRNA	linear	HTC 12-JUL-2001
LOCUS	BC005293	Homo sapiens, Similar to nibrin, clone IMAGE:3997534, mRNA.			
DEFINITION	BC005293				
ACCESSION	BC005293.1	GI:14710042			
VERSION	HTC				
KEYWORDS	human,				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1497)				
AUTHORS	Strausberg, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (27-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	Contact: MGC help desk Email: cgaps-femail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: CLONTECH Laboratories, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www-shgc.stanford.edu Contact: (Dickson, Mark) mcd@paxil.stanford.edu R. M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, D. M.				
FEATURES	Location/Qualifiers				
source	1. .1497				
	/organism="Homo sapiens"				

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 16 Row: j Column: 6
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 3098674
This clone has the following problem: frame shifted.

```
/db_xref="taxon:9606"
/clone="IMAGE:3997534"
/tissue_type="Bladder, carcinoma"
/clone_lib="NIH_MGC_53"
/lab_host="DH10B"
/notes="vector: pDNR-LIB"
BASE COUNT      530 a 282 c 317 g 368 t
ORIGIN

Query Match      33.2%  Score 1460.2;  DB 11;  Length 1497;
Best Local Similarity 99.4%;  Pred. No. 9.6e-212;
Matches 1476;  Conservative 0;  Mismatches 2;  Indels 1;  Gaps 1;

QY 22 acgtcgcccccagccctgagagcgccgaccgatgtgaaactgctgcccgcgcggggccc 81
DB 2  ACGTCGGCCCCAGCCCTGAGGAGCGGACCGATGTGGAACCTGCTGCCCGCGCGGCC 61

QY 82 ggcaggaggaaccatacacagaacttttgactgcggttgagtacgtttgttggaaagaaaa 141
DB 62 GCGAGGAGGAGAACCATACAGACTTTTGACTGGCGTTTGAGTACGTTGTTGGAAGGAAAA 121

QY 142 ctgtgcattcttaattgaaatgatcagtcgatcagccgaatcatgctgtgttaactgc 201
DB 122 CTGTGCCATCTTAATTGNAATGATCAGTCGATCAGCGGAATCATGCTGTGTTAACTGC 181

QY 202 taactttctgttaaccaacctgagtcacacagatgaatccctgtattgacattaaaaa 261
DB 182 TAACCTTTCTGTAACCAACCTGAGTCAACACAGATGAAATCCCTGTATTGACATTAAGA 241

QY 262 taattctaagtatggtacctttgtaagtgaagaaaaaatgagaaatggcctttcccgaa 321
DB 242 TAATTCTAAAGTATGGTACCTTTGTTAATGAGGAAAAAATGAGAAATGGCTTTTCCGAA 301

QY 322 ttgaagtcggggatggtattacttttgaggtgttggagtaaatcagaatagagta 381
DB 302 TTTGAAGTCGGGGATGGTATTACTTTTGGAGCTTTTGGAAAGTAAATTCAGAAATGAGTA 361

QY 382 tgagcctttggtgcagtcgtcttctgttagatgtctctgggaaaaactgcttaataca 441
DB 362 TGAGCCTTTGGTTCATGCTCTTCTGTTTATAGTGTCTCTGGGAAACTGCTTTAAATCA 421

QY 442 agctatttgcaacttgaggatttactgttaacaattggacagaagaatgacctacct 501
DB 422 AGCTATATTGCAACTTGGAGATTACTGTAAACAATTTGGACAGAAATGCATCACT 481

QY 502 tgtcatggtatcagtgaaagtaccattaaacaatatgtgcactcatttggcagctcc 561
DB 482 TGTCTGTTACGTGAAGATTACCATTAACAATAATATGTGCATCATTTGTGGCGTCC 541

QY 562 aattgtaagccagaatattttactgaattcctgaaagcagttcagttccaagaagcagcc 621
DB 542 AATTGTAAGCCAGCAATATTTACTGAATTCCTGAAAGCAGTTTCAGTCCAAAGAGCAGCC 601

QY 622 tccacaattgaaagtttttaccacacctcttgatgaaccatctatttggaaagtaaaatgt 681
DB 602 TCCACAATTTGAAAGTTTACCACACCTCTTGTATGAACCATCTATTGGAAGTAAAAATGT 661

QY 682 tgatctgcagagcgagaaagaaacaaatcttcaaggggaaaaacatttatatttt 741
DB 662 TGATCTGTACGACGCGGAGGAAAGAAACAAATCTTCAAGGGGAAACATTTATATTTT 721

QY 742 gaatgcaaacagcataagaattgagttccgcagttgtctttggaggtggggaagctag 801
DB 722 GAATGCCAAACAGCATAGAATTTGAGTTCGCGAGTTGTCTTTTGGAGTGGGGAAGCTAG 781

QY 802 gttgataacagaagaatgaagaagaacataaatttttttggctccgggaacgtgtgt 861
DB 782 GTTGATAACAGAGAATGAAGAAGAACATAAATTTCTTTTGGCTCCGGGAACGTGTGT 841

QY 862 tgtgtatcaggaataacaaactcacagaccttaattcctgactgtcagaagaatgat 921
DB 842 TGTGTATACAGGAATAACAACTCACAGACCTTAATTCCTGACTGTGTCAGAAAGAAATGGAT 901
```

```
QY 922 tcagtcaataatggtatgtctccaaagcaaggtcttagacctattctctgaagcagaat 981
DB 902 TCAGTCAATAATGATATGTCTCCAAAGGCAAGGCTTAGACCTATTCTCTGAAGCAGAAAT 961

QY 982 tggattggcggtgatcttttcatactgactacaagaataactgtatctctcaggccatccag 1041
DB 962 TGGATTGGCGGTGATTTTTCATGACTACAAAGAAATTTACTGTGATCCTCTCAGGGCCATCCAG 1021

QY 1042 tacaggataaagacaacaactccaggaccagcctttccaaagcgctgtcagttgatga 1101
DB 1022 TACAGGATTAAAGACAACAACCTCCAGGACCAGCCTTTCAAGGCGTGTCTCAGTTGATGA 1081

QY 1102 aaaactaatgccaagcgcccgagtgaaactacaactacatcagtagtgacacagaatcaga 1161
DB 1082 AAAACTAATGCAAGCGCCCGCAGTGAACACTACAACATACGTAGCTGACACAGAATCAGA 1141

QY 1162 gcaagcagatacatgggatttgagtgaaaggccaaagcaaaagaatactctcaaaatga 1221
DB 1142 GCAAGCAGATACATGGGATTTGAGTGAAGGCCAAAGAAATCAAAAGTCTCCAAATGGA 1201

QY 1222 acaaaaattcagaatgctttcacaaagcagcccccactgtaaaggagctcctgcaaaacaag 1281
DB 1202 ACAAATTCAGATGCTTTTCAAGAGCGCACCCACTGTAAGGAGTCTCTGCAAAACAAG 1261

QY 1282 ctctaaataatagtagtgatcaaatcactttggctaagatgagaatccccaaactatca 1341
DB 1262 CTCTAATAATAATAGTAGTGGTATCAAAATACCTTTGGCTAAGATAGAATCCCAAACTATCA 1321

QY 1342 gctttccaaactaaaattggccaagtataataaaaagtaaaagtagggcttctcagcaga 1401
DB 1322 GCTTTCCAACTAAATTTGCCAAGTATAAATAAAGTAAAGTAGGGCTTCTTCAGCAGCA 1381

QY 1402 gcagacaaactccatcagaaactacttccagcgtctacc-aaaaaaagggaaagggatg 1460
DB 1382 GCACCAACTCCATCAGAAACTACTTTCAGCCGCTTACCAAAACAAAAAGGAAAGGGATG 1441

QY 1461 aagaaatcagaagaatgcttctcatgcaaatcagcagaataagaaa 1505
DB 1442 AAGAAATCAAGAAATGCTTCATGTCACAAAAAAGGAAAAAAGGAAAAA 1486
```

RESULT 2

```
BC016762
LOCUS      1460 bp  mRNA  linear  HTC 09-NOV-2001
DEFINITION Homo sapiens, Similar to Nijmegen breakage syndrome 1 (nibrin),
clone IMAGE:4104186, mRNA.
ACCESSION BC016762
VERSION   BC016762.1 GI:16876977
KEYWORDS  HTC.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1460)
AUTHORS   Strausberg, R.
TITLE     Direct Submission
JOURNAL   Submitted (05-NOV-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
Email: gcaps-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 32 Row: e Column: 20
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6996019
This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
1..1460
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4104186"
/tissue_type="Brain, glioblastoma"
/clone_lib="NIH_MGC_57"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
BASE COUNT 517 a 269 c 313 g 361 t
ORIGIN

Query Match 32.4%; Score 1427; DB 11; Length 1460;
Best Local Similarity 99.3%; Pred. No. 1e-206;
Matches 1433; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 33 agccctgagagcgaccgatgtgaaactgtgcccgcgcggccgcggcgagag 92
DB 5 AGCCCTGAGGAGCGGACCGCATGTGGAACCTGCTGCCGCGCGCGCGCAGGAG 64
QY 93 aaccatcacagactttgactgcggtgagtcagctgttggaagaaaaactgtgccattc 152
DB 65 AACCATACAGACTTTTGACTTGGCGTGTGAGTACGTGTGTGGAAGAAAACTGTGCCATTC 124
QY 153 taattgaaatgatcatgctgacgcgaaatcatgctgtgttaactgtcaactttctg 212
DB 125 TGATTGAAATGATCATGCTGATCAGCGGAATCATGCTGTGTAACTGCTAACTTTCTG 184
QY 213 taaccaactgagtcgaacagatgaatccctgtattgacattaaagataattcttaagt 272
DB 185 TAACCAACCTGAGTCAACAGATGAATCCCTGTATTGACATTAAGAGATAATCTAAGT 244
QY 273 atggtaccttgttaatgagaaaaaatgcagaatgcttcccgaaacttgaagtcgg 332
DB 245 ATGCTACCTTTGTATGAGGAAAAAATGCAGATGCTTTTCCCGAACCTTGAAGTCGG 304
QY 333 gggatggtattacttttggagtggtttggaagtataatcagaatagatgagcctttg 392
DB 305 GGGATGCTATTACTTTTGGAGTGTGTGGAAGTAAATTCAGAATAGATGATGAGCCTTTGG 364
QY 393 ttgcatgctcttctgttttagatgtctctgggaaactgctttaaatacaagctatatgc 452
DB 365 TTGCATGCTCTCTTGTGTTAGATGTCTCTGGGAAACTGCTTTTAAATCAAGCTATATTCG 424
QY 453 aacttgaggatttactgtaaacaattggacagaagaatgcactcaccttgtcatggtat 512
DB 425 AACTTGGAGATTACTGTAAACAATTTGGACAGAAGATGCACCTACCTTGTCAATGATAT 484
QY 513 cagtgaagttaaccataaacaatatgtgactcatcttggacgctccaattgtaagc 572
DB 485 CAGTGAAGTTACCATTTAAACAATATGTGCACTCACTTTTGGACGTCCTCAATTTGAAAGC 544
QY 573 cagaatatttactgaattcctgaagcagttcagtcocaaagaagcagcctcccaaatg 632
DB 545 CAGAAATATTTACTGAATTCCTGAAAGCAGTTGAGTCCAAAGAAGCAGCCTCCCAAAATG 604
QY 633 aaagtgtttaccacactcttgatgaaccatctattggaagtaaaaaattgtgactgtcag 692
DB 605 AAAGTTTAAACCACTCTTGTATGAAACCATCTATTGGAAGTAAATTTGATCTGTGAC 664
QY 693 gacggcaggaagaaacaaactctcaaaagggaacatttatatttttgaatgccaaac 752
DB 665 GACGGCAGGAAGAAAAACAAATCTTCAAGAGGAAAAACATTTATATTTTGAATGCCAAC 724
QY 753 agcataagaattgagttccgcggtgtcttcttgagggtgggagctgagttgataacag 812

DB 725 AGCATAAAGAAATTGAGTTCGCCAGTTCTCTTTGGAGGTGGGAAAGCTAGTTGATAACAG 784
QY 813 aaagaatgaagaacataaattcttttggctccggaacgtgtgtgtgtatcacag 872
DB 785 AAGAAATGAAGAGAACAATAATTTCTTTTGGCTCCGGAACCTGTGTGTGTATACAG 844
QY 873 gaataacaaactcacagacaccttaattcctgactgtcagaagaaatggattcagtcataa 932
DB 845 GAATAACAAACTCACAGACCTTAATCTGACTCTCAGAAGAAATGGATTGAGTCAATAA 904
QY 933 tggatatgctccaaaggaaggtcttagaacctattcctgaagcagaagaattggattggcgg 992
DB 905 TGGATATGCTCCAAAGCAAGGCTTTAGACCTATTCTGAAAGCAAAATTTGGATTGGCGG 964
QY 993 tgattttcatgactacaaagaattactgtgactcctcagggccatcccgactacagagataa 1052
DB 965 TGATTTTCATGACTACAAAGAAATTTACTGTGATCTCTCAGGCGCATCCCAAGTACAGATTA 1024
QY 1053 agacaacaactccaggaccacgacctttcacaaaggcgtgtcagttgtatgaaacaaactatgc 1112
DB 1025 AGACAACAACCTCCAGGACCAAGCCTTTTCAAGGCGGTGTCTAGTTGATGAAAACTAATGC 1084
QY 1113 caagcgcgccagtgaaacactacatacagtagctgacacagaatacagagcaagacata 1172
DB 1085 CAAAGCGCCAGTGAACACTACAACTACGTAGCTGACAGAAATCAGAGCAAGCAGATA 1144
QY 1173 catgggatttgagtgaaagcgaagaaatcacaagctcccaaatggaacaaattca 1232
DB 1145 CATGGGATTTGAGTGAAGGCCAAAGAAATCAAAAGTCTCCAAATGGACAAANATCA 1204
QY 1233 gaatgctttcacagacgcacccactgtlaaaggagtcctgcacaaacagctcttaataa 1292
DB 1205 GAATGCTTTCAAGATGTCACCCACTGTAAAGGAGTCTGCAAAACAAGCTCTAATAATA 1264
QY 1293 atagtaggtatcaaatactttggcctaagatgagaatcccaaatatcagctttccacca 1352
DB 1265 ATAGTATGGTATCAAAATACTTTGGCTAAGATGAGAAATCCCAACTATCAGCTTTTCACCA 1324
QY 1353 ctaaattgcgaagtataataaaagtaagataggctcttcagcagcagcagaccact 1412
DB 1325 CTAATTTGCCAAGTATTAATAAAGTAAGATAGGGCTTCTTAAGCAGCAGCAGACCAACT 1384
QY 1413 ccatacagaactactttcagcctgtcccaaaaaaggggaagggatgaagaaatacaag 1472
DB 1385 CCATCAGAAACTACTTTTCCAGCCGTCTACCAAAAAAGGGAAGGAAAAAATAAAAAA 1444
QY 1473 aaa 1475
DB 1445 AAA 1447

RESULT 3
BC003719
LOCUS BC003719 2518 bp mRNA linear HTC 12-JUL-2001
DEFINITION Mus musculus, nibrin, clone IMAGE:3485295, mRNA.
ACCESSION BC003719
VERSION BC003719.1 GI:14708252
KEYWORDS HTC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2518)
Strausberg, R.
Direct Submission
Submitted (28-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I. M. A. G. E. Consortium (LINL)
 DNA Sequencing by: Sequencing Group at the Stanford Human
 Genome Center, Stanford University School of Medicine, Stanford,
 CA 94305
 Web site: <http://www.shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 7 Row: e Column: 23
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 3676834
This clone has the following problem: frame shifted.

FEATURES

source

```

/organism="Mus musculus"
/db_xref="locusID:27354"
/db_xref="taxon:10090"
/clone="IMAGE:3485295"
/tissue_type="Mammary tumor
old, gross tissue."
/clone_lib="NCI CGAP_Mam5"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
a 527 c 583 g 562 t

```

BASE COUNT
ORIGIN

Query Match 31.6%; Score 1391; DB 11; Length 2518;
Best Local Similarity 75.0%; Pred. No. 2.5e-201;
Matches 182; Conservative 0; Mismatches 590; Indels 17; Gaps 6;

Qy	37	ctgagagccgcgaccgatgtggaaactgctgcccgccgcccggcgcaggagagagaacc	96
Db	103	CTGAGACCGCGCCCATGTGGAAAGCTGCTCCGCGCGCGGTGC	162
Qy	97	atacaacatttgactggcgttgaagtacatttgtgaagaaaacttgccatctcta	156
Db	163	ATACCGACTTTTGGCGCGGTGAGTACGTGTGGAGGAATACTGGGCATTCTGAT	222
Qy	157	tgaataatgatcagtcagtcagcgcgaatacatgctgtgttaactgctaaactttctgttaac	216
Db	223	TGAAAATGATCAGTCATCAGTCGAACCACCTGCTGCTTAACAGTAAACTTCCCTCTAAC	282
Qy	217	caacctggagtcacaacagatgaaatccctgttatgtacattaaagataaatcttaagtatgg	276
Db	283	CAGTTTGAGTCAACACAGATGAAATTCCTACATTAACAATAAAGAATAATCTAAGTATGG	342
Qy	277	tacctttgtaatgagaaaaaatgcagaatggcttttccgaactttgaagtcggggga	336
Db	343	AACCTTTGTTAATGAGAAAAATGCAGACTGGTCTTCTCGCACGTTGAAGACAGGAGA	402
Qy	337	tggttactcttttggagtgcttggaaagtaaattcagaaatagagatatggaccccttggttgc	396
Db	403	TAGAGTTACCTTTGGGGTGTTCAAAGTAAATTCAGAGTAGAATACGACGCCCTTGGTGTCT	462
Qy	397	atgctctctctgttagatgtctctgggaaaaactgctttaaatcaaagctatatattgcaact	456
Db	463	TTGTTCTTCTTGTTTTAGATGCTCTCGGAAAACTGTTTTAAATCAACAGCTATTTTACAGCT	522
Qy	457	tgagagatttactgtaaacaattggacagaagaatgcactcaccttgcctatggtatcagt	516
Db	523	TGAGGAGACTTACCGCAAAACCTGGACGGAAGAATGTACCCACTTGTCTATGTGCGCAGT	582
Qy	517	gaaagtaccacttaaaacaatatgtgcactcatttggagcgtccaattgttaagccaga	576
Db	583	TAAAGTCACCATTAAACATATATGTGCACCTCATTTTGGAGCTCCAATTTATAAACCCAGA	642
Qy	577	atatattactgaattoctgaaagcagttcagttccaagaagcagctcccacaattggaag	636
Db	643	ATACTTTCTTGAATTTCTCAAGACAGTTGAAATCTAAGAACAGCCCTCCAGATATTGAAAG	702

5
6
7
8
9

[illegible]

```

QY 1714 tgaatggtgcatgaagaatgaagtattggacacagttattccaaggacacaaacacagagtt 1773
Db 1772 TGATTAGCTTACAGAGAGAGAGCTTCTTGAAGAGTACTCAGGAGTACAAAGCCAGAGTT 1831
QY 1774 agaaattgatgtgaagttcaaaaacagaggaagatgtcaattgtagaaaaagggccaag 1833
Db 1832 GGCAGTTCAAGTGAAGGTTGAAAGCAGAGGAGGAGATGACACCATCATGAAAAAGCCAAAG 1891
QY 1834 gatggaatagaacaaatgacactttcagttcagtgatgaagcagtgaccagaagaatgagcaaat 1893
Db 1892 GATGGACGACAGAGGAAATCGGCCCTGTAATGGTGGATCAGAGCCGGAAGACACAGCGC 1951
QY 1894 atctcaagaataatgaattgggaagaaacgtgaactcaaggaagactcaactatggtcagc 1953
Db 1952 TCTTCAAGAAAGATGAAGAAAGAAAGAAAGATGAATTCACACAGAGTCGTGGTCAACAAA 2011
QY 1954 taagaataatctaaacatgaacaaacttcaggtatgatgatgatgagatgcttcccaaaaaagct 2013
Db 2012 ACATGAATAAGCTTAATAGTATGATGCTCTCAGGACAGCAGTGAGGAGTGCACCGGAACCT 2071
QY 2014 gttattgactgaattagatcactggtgattaaaaactctacttccagaaatccgctcgg 2073
Db 2072 GCTGCTGACTGAATTAAGTTCATTTGGTGTGTCAGTACCAACCACTCCACTCCAGAAATCT 2131
QY 2074 cataatgatgatgttgatcaactaaaaatttcaagaaatttcaaaaaggtcacatctc 2133
Db 2132 ATGTGTAATGAATGTGTGCTCACTGAAGAAATTTCAAGAAATTTCAAGAAAGCGGACATTTCC 2191
QY 2134 tgaagcaggaataacttccacacatcattggagatcagatcactcaatgaactcattcgtcg 2193
Db 2192 TGGGGCAGGAAGACTTCCACACATCATCGGAGGATCAGATTTAGTAGGCCACCATGCTCG 2251
QY 2194 aaagaatcacagaactgaagagtggttaagcaggaataatgaggtacaaaaatcaaatcgc 2253
Db 2252 AAAGAATACCGAGTTAGAGAAATGTTGAAACAGGAAATGGAGGTACAGAAACACAAAGC 2311
QY 2254 aaaaagaagctctctgctgatgatcttttagatacaatccctatttaaaagggagaag 2313
Db 2312 AAAGGAGGAATCTCTTCTGATGATCTGTTTAGATATATATCTTAATGTAATAA---AGAAG 2368
QY 2314 ataactgaggtatttaaaagaagccatggaataactctcctagtaagcatctacttcagg 2373
Db 2369 ATAAATGAACCTTGTAAAGGAAGCTGTAGAAACACATTTCTACTTCAGACCAACTGAGAT 2428
QY 2374 ccacaaggttatataatataatgtatgataagaagcatttaagttacaaatggtttatg 2433
Db 2429 ACATTAATATAGTACATAGAAATTAATAATGCTTTAAGGTTTCAGTTTATTAATAAAGG 2488
QY 2434 cctaaatttataataataaatagcacaaaa 2462
Db 2489 AAACAACACTCTAAAAAATAAAAAAAAAA 2517

RESULT 4
LOCUS BC013190 1792 bp mRNA linear HTC 29-AUG-2001
DEFINITION Mus musculus, Similar to nibrin, clone IMAGE:3602503, mRNA.
ACCESSION BC013190
VERSION BC013190.1 GI:15341986
KEYWORDS HTC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1792)
Strausberg,R.
Direct Submission
Submitted (27-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
```

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
CDNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalon@bcm.tmc.edu.
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 17 Row: 9 Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 3676834
This clone has the following problem: incomplete processing.

FEATURES
Location/Qualifiers
1..1792
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3602503"
/tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
ductal carcinoma, 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

BASE COUNT 593 a 394 c 404 g 401 t

ORIGIN

Query Match 24.8%; Score 1090.4; DB 11; Length 1792;
Best Local Similarity 79.4%; Pred. No. 8.7e-156;
Matches 1331; Conservative 0; Mismatches 336; Indels 9; Gaps 3;

QY 37 ctgaggagccggacccgcatgtggaaaactgtcccgccgcccggccgcgagagaggaacc 96
Db 92 CTGAGAGCCGCCGCCCATGTGGAAGCTGCTCCCGCCGCGGTGCGGTCCAGGAGAACC 151
QY 97 atacagactttgactggcgttgagtagcgtgtgttggaaggaagaaactgtgccattcatt 156
Db 152 ATACCGACTTTTGGCCGCGGTGAGTACGTGTTGGGAGGAAACAACTGTGCAATTCGAT 211
QY 157 tgaataatgacagtcagtcagcgaataatcgtctgttaactcacttaactttctgtaac 216
Db 212 TGAATAATGATCAGTCAATCAGTCGAAACCATGCTGCTTTAACACATAAACTTCCCTGTAAC 271
QY 217 caacctgagtcacacagatgaaatccctgtattgacattaaaaagataattcctaagtgg 276
Db 272 CAGTTTCAGTCAACACAGATGAAATTCCTACATTAAACAATAAAGATAATTTCTAAGTATGG 331
QY 277 tacctttgtaataaggaaaaaatgcagaatggcctttcccgaaactttgaagtcggggga 336
Db 332 AACCTTTGTTAATGAAGAAAAAATGCAGACTGGTCTTTTCTGTCACGTTGAAGACAGAGA 391
QY 337 tggattacttttggagtggtttggaagttaattcagaagtagatagcgttccttggttc 396
Db 392 TAGAGTTACCTTTGGGGTGTGTTGAAGTAAATTCAGAGTAGAATACGAGCCCTTGGTGT 451
QY 397 atgctctcttctgttagatgtctctgggaaaaactgctttaaatacaagctatttgcac 456
Db 452 TTGTTCTCTCTTGTGTAGATGCTCTCGGAAAAACCTGTTTAAATCAAGCTATTTTACAGCT 511
QY 457 tggaggattactgtataacaaatggacagaagaatgcactcactccttgtcatggtatcagt 516
Db 512 TGGAGGACTTATACCCAAACAACTGGACGGAAGAAATGTACCCACCTTGTTCATGTCGGCAGT 571
QY 517 gaaagtaccattaaaaaataatgtgcactcatttggagcgtcccaattttaaagccaga 576
Db 572 TAAAGTCACCATTAAAACTATATGTGCACTCATTTTGGAGCTCCCAATTAATAAACGAGA 631


```
Db 321 TTTGGCTCCGGGACGCTGTCTTGTGATACAGGAATAACAAACTCACAGACCTTAATTC 380
QY 901 tgactgtcagaataatggattcagtcataatgaatgatatgtccaaaggaagcttag 960
Db 381 TGACTGTGAGAAGAAATGGGATTCAGTCAATAATGGATATGCTCCAAAGGCAAGCTCTAG 440
QY 961 acctattcctgaagcagaataatggattggcggtgatttcatgactacaagaattactg 1020
Db 441 ACCTATTCTTGAAGCAGAAATGGATTGGGTGGGTTGATTATCATGACTACAAAGAAATTA 500
QY 1021 tgatcctcaggggccatccccagtcacaggattaaagacaaactccaggaccacgcttc 1080
Db 501 TGAATCTTCAGGGCCATCCCATGATACAGGATTAAAGACAACTCCAGGACCAAGCCCTTC 560
QY 1081 acaaggcgtgtcagtgatgaaaaactaatgccaagcgccccagtggaacactacaacata 1140
Db 561 ACAGGGCTGTGATGATGAAGAACTAATGCCAAGCGCCCATGAGTGAACACTACAACATA 620
QY 1141 cgtagctgacacagaatcagacgaagcagatatacatggtgatttgagtgaaaggccaaaga 1200
Db 621 CGTAGCTGACACAGAATCAGACGAAGCAGATACATGGGATTTGAGTGAAAGGCCAAAGA 680
QY 1201 aatcaaatgtccaaatggaacaaataatcagaatgcttttcacaaagcaccaccactgt 1260
Db 681 AATCAAGTCCCAAAATGGAACAAAAATTCAGAAATGCTTTTCACAAAGCACCACCTGT 740
QY 1261 aaaggagctctgcaaaacagctctaataataatagatgatatgatacttctggctaa 1320
Db 741 NAAGGAGCTCTGCANATAAGCTCTAAT-ATAATAGTATGGGATCAAAATCTCTGGGTAA 799
QY 1321 gatgagaatcccaactcagctttccaccaactaaattgccagtagtaataaagtaa 1380
Db 800 GATGAGAATNCCAACTATCAGCTTTCCACCACTTAATTGCCCAAGTATTAATFAAAGTA 859
QY 1381 agatagggtcttcagcagcagcagacccaactccatcagaaactacttcagcogctac 1440
Db 860 AGATAGGGCTCTCAGCAGCAGAGAACCAACTCATCAGAAACCTTCCAGGCGCTTAC 919
QY 1441 caaaaa 1447
Db 920 CACAAA 926

RESULT 6
LOCUS BG214621 789 bp mRNA linear EST 21-APR-2001
DEFINITION RST34264 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG214621
VERSION BG214621.1 GI:13740642
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 789)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
E., Veloso,N., Klika,A., McElligott,K., Booser,S., Mays,R., Smith
J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
JOURNAL 2127151
MEDLINE Contact: Scott J. Cain
COMMENT Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scai@atersys.com
High quality sequence stop: 545.
Location/Qualifiers
```

```
source 1..789
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 284 a 156 c 160 g 189 t
ORIGIN

Query Match 16.8% Score 741: DB 10: Length 789;
Best Local Similarity 97.8% Pred. No. 8.3e-103;
Matches 772; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

QY 637 tttttacccacctcttgatgaac-cattctattggaagtataaatgttgatctgcaggac 695
Db 2 TTTTATTACCACCTCTTGTGATGAACATCATCTATTGGAAGTAAAAATGTTGATCTGTCA 61
QY 696 ggcaggaaagaaacaaatcttcaaaagggaacattttatttttgaatgccaaacagc 755
Db 62 GGCAGGAAGAAAAACAATCTTCAAGGGGAAAAACATTTATATTTTGAATGCCAAACAGC 121
QY 756 ataagaattgagttccgcaggtgtcttcttgaggtgggaagctaggttgataacagaag 815
Db 122 ATAAAGAAATTTAGTTCCTCCAGTGTCTCTTTGGAGGTGGGAGGCTAGGTTGATATA 181
QY 816 agaatgaagaagaacataattcttttgcctccgggaacgtgtgtgttgatcacagaa 875
Db 182 AGATGAAGAAGAACATAATTTCTTTTGGCTCCGGGAACGCTGTGTGTTGATACAGGA 241
QY 876 taacaaactcacagacetttaattcctgactgtcagaagaaatggattcagtcataatgg 935
Db 242 TAACAAACTCACAGACC-TGATTCTCTACTGTCAGAGAAATGGATTTCAGTCAATAAT 300
QY 936 atatgtctccaaaggcaaggctcttagacctattcctgaagcagaataatggattgcggtga 995
Db 301 ATATGCTCCAAAGGCAAGGCTTTAGACCTATTCTCTGAAGCATATAATTTGGATTGG 360
QY 996 ttttcagtactacaaagaattactgtatcctcaggggccatcccagtlacaggattaaaga 1055
Db 361 TTTTCATGACTACAAAGAAATTTACTGTGATCCTCAGGGCCATCCCAGTACAGGATTA 420
QY 1056 caacaactccaggaccagcctttcacaaagcggtgtcagttgataaaactaatgccaa 1115
Db 421 CAACAACCTCCAGGACCAGGCGCTTTCACAAGGCGTGTCTGATGATGAAAAAACTAAT 480
QY 1116 gcgccccagtgaaacactacatacatagctagctgacacagaatacagacagacatacat 1175
Db 481 GCGCCCCAGTGAACACATACAACTACATACGTAGCTGACAGAAATCAGCCGACGATACAT 540
QY 1176 ggaatttgagtgaaagcgccaaagaatacaaaagtctccaaatggacaaataatccagaa 1235
Db 541 GGGATTTGACGAAGGCGCCAAAGAAATCAAAAGTCTCCAAATGGAACAAATAATTCAGAA 600
QY 1236 tgccttcacaagaagcagccaccactgttaaaggagtcctgtcgaacaaagctcttaataata 1295
Db 601 TGCTTTTCAAGAGCGCACCCACCTGCTAAAGGAGTCTCTCAAAACAAAGCTCTAATAATA 660
QY 1296 gtatggtatacaatacttttggttaagatgagaatcccaactatcagctttccaccaacta 1355
Db 661 GCATGGTATCAAAATACCTTTGACTAAGATGAGAAATCCCATCTACTCATCTTTTCACCA 720
QY 1356 aattgccaagtataataaagaataagtagggcttctcagcagacagacagacccaactcca 1415
Db 721 AATTGCCCAAGTATAAAATAAAGTAATGATAGGGCTTCTTCATCAGCAGCAGACCAACTCCA 780
QY 1416 tcagaaact 1424
||||| |||
```



```
Db 600 AATGAAATCGGGGAAGAAACGTGAACCTCAAGGAAGACTCACTATGGTCAGCTAAAGAAAT 659
Qy 1963 atctaaacaatgacaaacttcaggatgatagtagatcttcc-aaaaaacgtgtattga 2021
|||||
Db 660 ATCTAACANTGACAACTTCAGGATGATAGTAGATCTTCCAAAAAAGCTGTCATCGA 719
Qy 2022 ctgaa-tttagatcactggtgattaaaaa 2049
|||||
Db 720 CTGAATTTAGATCACTGGTGATTAAAAA 748

RESULT 10
BG284646
LOCUS 602408824F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4537973 5',
DEFINITION mRNA sequence.
ACCESSION BG284646
VERSION BG284646.1 GI:13035811
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 960)
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10463 row: p column: 06
High quality sequence stop: 771.
Location/Qualifiers
1..960
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4537973"
/clone_lib="NIH_MGC_91"
/tissue_type="adenocarcinoma, cell line"
/notes="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 328 a 185 c 220 g 227 t
ORIGIN

Query Match 15.8%; Score 696.2; DB 10; Length 960;
Best Local Similarity 99.1%; Pred. No. 4.8e-96;
Matches 742; Conservative 0; Mismatches 3; Indels 4; Gaps 4;

Qy 517 gaaagttaccattaaaacaatatgtgactcatttggagcgtccaattgtaaaagccaga 576
Db 1 GAAAGTTACCATTAAAACAATATGTGCACCTATTTGTGGAGCTCCAA-TGTAAAGCCAGA 59
Qy 577 atatttactgaattctgaaagcagttcagtcacgaagcagcctccacaattgaaag 636
Db 60 ATATTTACTGAATTCCTGAAGCAGTTTCAGTCCCAAGAGCAGCTCCACAAATTGAAG 119
Qy 637 ttttaccacaccttggatgaaccattctattggaagtaaaatgttgatctgtcaggacg 696
Db 120 TTTTACCACCTCTTGTATGAACCATCTA-TGGAAGTAAAAATGTTGATCTGTCCAGGACG 178
Qy 697 gcaggagaagaaacaaatcttcaagggaagaaacatttatattttgtaatgcacaacagca 756
|||||
```

```
Db 179 GCAGGAAGAAAAACAATCTTTCAAGGGAAAAACAATTTATATTTTGAATGCCAAACACCA 238
Qy 757 taagaaattgagttccgcagttgcttcttggaggtggggaagctaggttgataacagaaga 816
|||||
Db 239 TAAGAAATTGAGTTCCGACGTTGCTTTGGAGGTGGGAAGCTAGGTTGATAACAGAAGA 298
Qy 817 gaatgaagaagaacataaattcttttggctccgggaacgtgtgttctttgtatcacagaat 876
|||||
Db 299 GAATGAAGAAGAACATAAATTTCTTTTGGTCCGGGAACGCTGTCTTGTGTATACAGGAAT 358
Qy 877 acaaaactcacagaccttaattctctgactgtcagaagaataatggattcagtcataatgga 936
|||||
Db 359 AACAAATCTCAGACACCTTAATTCCTGACTGTCTCAGAAGAAATGGATTTCAGTCAATAATGGA 418
Qy 937 tatgtccaaaggaaggtcttagacattctcctgaagcagaataatggattggcgtgat 996
Db 419 TATGCTCCAAAGGCAAGGCTTAGACCTATTCTTGAAGCAGAAATTTGGATTGGCGGTGAT 478
Qy 997 tttcatgactacaaaagaattactgtgactcctcagggccatccccagtcacagattaaagac 1056
|||||
Db 479 TTTTCATGACTACAAAGAAATTAATCTGTGATCTCTCAGGGCATCCCACTACAGGATTAAAGAC 538
Qy 1057 acaaaactccaggaccacccctttcacagcgctgtcagttgatgaaaaactaatgccaaag 1116
|||||
Db 539 AACAACTCCAGGACCAAGCCTTTCCAAAGCGGTGTGAGTTGATGAAAACTAAATGCCAAG 598
Qy 1117 cgccccagtgaaactacacatactagctgacacagaatcagacgaagcaagcagatcacatg 1176
Db 599 CG-CCCAGTGAACACTACACATACGTAGCTGACACAGAATCAGACGAGCAAGCAGATACATG 657
Qy 1177 ggatttgagtgaaaggccaaaagaaatcaaaagtctcaaaatggaacaaattcagaat 1236
Db 658 GGATTTGAGTGAAAGG-CAAAAGAAATCAAAAGTCTCCAACTTGGACAAACAAAATTCAGAA 716
Qy 1237 gctttcacagacccacccactgtaaaagg 1265
Db 717 GCTTTCACAGAGCCACCCACTGTGAAGG 745

RESULT 11
BG388866
LOCUS 602414591F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4522774 5',
DEFINITION mRNA sequence.
ACCESSION BG388866.
VERSION BG388866.1 GI:13282312
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 779)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10424 row: f column: 23
High quality sequence stop: 722.
Location/Qualifiers
1..779
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4522774"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"

FEATURES
source
```

/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-qt primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
249 a 126 c 174 g 230 t

BASE COUNT
ORIGIN

Query Match 15.7%; Score 690.8; DB 10; Length 779;
Best Local Similarity 96.9%; Pred. No. 3.3e-95;
Matches 758; Conservative 0; Mismatches 17; Indels 7; Gaps 5;

QY 260 gataattcgaatgacgttctgttaagagagaaatgcagaatgctttcccca 319
|||||
Db 1 GATAATCTAAGTATGGTACCTTGTGTAAGGAAATAATGCAGATGCTTTCCCGA 60
|||||

QY 320 acttgaagtcg9ggatgtattacttttggagtggttggaaagtaaatcagaatagag 379
|||||
Db 61 ACTTTGAAGTCGGGGATGCTATTACTTTTGGAGTGTTCGGAAGTAAATTCAGAATAGAG 120
|||||

QY 380 tatgagccttgggtgcatgctctctgtttagatgctctctgggaaactgctttaa 439
|||||
Db 121 TATGAGCCTTTGGTTGCATGCTCTCTTGTGTTAGATGCTCTCGGAAACTGCTTTAAAT 180
|||||

QY 440 caagctatatgcaacttgagattactgttaacaaatgtggacagagaatgcactcac 499
|||||
Db 181 CAAGCTATATGCACTTGAGAGATTTACTGTAACAAATTTGGACAGAAAGATGCATCAC 240
|||||

QY 500 ctgtcatggtatcagtgaaagtaccattataacaaatgtgtgcactcatttggagcgt 559
|||||
Db 241 CTGTGCTATGATCAGTGAAAGTTACCATTAAAAACAATATGTGCACCTCAITTTGTGGAGCT 300
|||||

QY 560 ccaatgtgaagccagaatattttactgaattctctgaaagcagctcgctccaagaagcag 619
|||||
Db 301 CCAATTGTAAGCCAGAAATATTTTACTGAAATTCCTGAAAGCAGTTGAGTCAAGAACAG 360
|||||

QY 620 cctccaaaattgaaagtctttaccacctcttgatgaaccatctattggaagtataat 679
|||||
Db 361 CTCCCAAAATTGNAAGTTTTTACCACCTCTTGTGATGAACCACTATTTGGAAGTAAAT 420
|||||

QY 680 gttgatctgcaggagcggcaggaagaaacaaatcttcaaaagggaacacatttatatt 739
|||||
Db 421 GTTGATCTGTGAGCAGCGCAGGAAGAAACAAATCTTCAAGGGAAACATTTATATTT 480
|||||

QY 740 ttgaatgccaaacagataagaattgagttccgcagttgtcttcttgagggtgggaaagct 799
|||||
Db 481 TTGAATGCCAACAGCATAAGAAATTTGAGTTCCGCACTTGTCTTTGGAGGTGGGGAAGCT 540
|||||

QY 800 aggtgtataacaga-agagaatgaagaagacaataattcttttggctccgggaacgtg 858
|||||
Db 541 AGGTTGTTACACACAGAGAAATGNAGNAGAACATAATTTCTTTTGGCTCCGGGAACGTG 600
|||||

QY 859 tgttgtgtacaggaataacaaactcacagacc-ttaattcctgactgtcagaagaat 917
|||||
Db 601 TGTGTTGTATACAGGAATAACAACTCACAGACCTTTAATTCCTGACTGTGTCAGAGAAT 660
|||||

QY 918 ggattcagtcataatggaatgctccaaaggcaagc-ttagaccattctctgaagca 976
|||||
Db 661 GGATTGAGTCAATGATGATGCTCCAAAGGCAAGGCTTTTAGACCTATTCTCTGAAGCA 720
|||||

QY 977 gaaattgattggcgggtgttttcagctacacaaagaattactgtgatctctcagggccat 1036
|||||
Db 721 GAA--TTGGATTGGGGGATTTTCATGACTACAA--GAATTCCTGGATCTCTCAGGGCAAT 776
|||||

QY 1037 cc 1038
||
Db 777 CC 778

RESULT 12
AT196269/c

LOCUS
DEFINITION
wh44q10.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2383650 3',
similar to TR:060672 060672 CELL CYCLE REGULATORY PROTEIN P95. [1]
// mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AT196269
AT196269.1 GI:5361732
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 712)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TUMOR
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps@femail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LENL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert length: 1177 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 451.
Location/Qualifiers
1. .712
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2383650"
/lab_host="NCI-CGAP_Kid11"
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

FEATURES
source

BASE COUNT 154 a 127 c 148 g 281 t 2 others
ORIGIN

Query Match 15.6%; Score 687.8; DB 9; Length 712;
Best Local Similarity 98.0%; Pred. No. 9.6e-95;
Matches 695; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 987 tggcgggtgatttctgactacaaagaattactgtgatctctcagggcctccccagtag 1046
|||||
Db 711 TTGCGGTGATTTTCTGACTACAAAGAATACCTGCTGATCTNCAGGGCCATCNCAGTACAG 652
|||||

QY 1047 gattaagacaacaactccaggaccagccttccacaagcgctgcagttgatgaaaaa 1106
|||||
Db 651 GATTAAAGACAACAACATCCAGGACCAAGCCCTTTTCACAGGCGTGTCTGATTTGATGAAAAAC 592
|||||

QY 1107 taatgccaaagccccagtgaaactacaactacatagctgacacagaatccagacaag 1166
|||||
Db 591 TAATGCCAAGCGTCCAGTGAACACTACACATACGTAGCTGACACAGAAATCAGAGCAAG 532
|||||

QY 1167 cagatacatgggatttgagtgaagggccaaagaaatcaaaagtcctccaaatggaaacaa 1226
|||||
Db 531 CAGATACATGGATCTCAGTGAAAGGCCAAAGAAATCAAAAGTCTCCAAAATGGACAAA 472
|||||

QY 1227 aattcagaatgctttccaaagacacccactgtaaggagtgctgcaaaacaaagctcta 1286
|||||
Db 471 AATTGCAATGCTTTTCAAGACGCACCCACTGTAAAGGAGTCTCTGCAAAACAAGCTCTA 412
|||||

```
Qy 1287 ataataatagtagtgcataacttttgctgaagatgagaaatcccaaaactatcagcttt 1346
Db 411 ATAATAATAGTAGTGTATCAAAATACTTTGGCTAAGATGAGAATCCCAAACATATCAGCTTT 352
Qy 1347 caccactaaattgccaaagtataataaaaaagtaagaatagggcttctcagcagcagaga 1406
Db 351 CACCAACTAAATTGCCAAGTATAATAAAAGTAAAGATAGGGCTTCTCAGCAGCAGCAGA 292
Qy 1407 ccaactcatcagaactacttccagcgtctacaaaaaggaaggaaggaatgaagaaa 1466
Db 291 CCAACTCATCAGAACTACTTTTCAGCGCTCTACAAAAGGAAGGAAGGATGAAGAAA 232
Qy 1467 atcaagaaatctctcatgcaaatcagcaagaatagaacgtcttctctcttttagaac 1526
Db 231 ATCAAGAAATGCTTCATGCAAAATCAGCAAGAATAGAAACGCTCTTGTTCTCTTTAGAAC 172
Qy 1527 aacacaaactgtcacacctcatcttggaataatagaagcagcagcatctatcagaaatg 1586
Db 171 AACACAACTGCTACACCTTCATTTGTGAAAAATAAGGAGCAGCATCTATCTGAGAATG 112
Qy 1587 agcctgtggacacaaactcagacaataacttattacagatacacagatttaaaatctattg 1646
Db 111 ACCCTGTGGACACAACTCAGACAATACACTTATTATTACAGATACAGATTTAAATCTATTG 52
Qy 1647 tgaataattctgcagtaaaatctcatctgcgcagaaaagctaaagctacaaa 1695
Db 51 TGAATAATTTCTGCCAGTAAATCTCATGCTGCAGAAAAGAAAGATACAA 3

RESULT 13
BG779728
LOCUS 602668556F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4808316 5',
DEFINITION mRNA sequence.
ACCESSION BG779728
VERSION BG779728.1 GI:14050045
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 735)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabps-femail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1660 row: p column: 13
High quality sequence stop: 735.
Location/Qualifiers
1..735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4808316"
/clone_lib="NIH_MGC_60"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfII (ggccgctcgcc); Site_2: SfII (ggccattatggcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGCGCCGACATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
```

```
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."
BASE COUNT 219 a 123 c 164 g 229 t
ORIGIN
Query Match 15.4%; Score 676.2; DB 10; Length 735;
Best Local Similarity 98.5%; Pred. No. 5.4e-93;
Matches 722; Conservative 2; Mismatches 5; Indels 4; Gaps 4;
Qy 2924 gtaccctttttgtggtcttactcagcgtgggtggtctgttttttgagctatttagag 2983
Db 1 GTACCCCTTTTGGTGGCTCTTACTCAGCTGGGTGGTCTCG-TTTGAGCTATTATTAGAG 59
Qy 2984 tcttagtttctacttataaagtagaaatgtgagattgttttttttttttttttttttt 3043
Db 60 TCCTAGTCTTTCCTATTATAAAGTAGAAATGGTGAGATGTTTTCTTTTTCTACCTTAAA 119
Qy 3044 gggagatggttaagaaacaaatgaatgtcttttttcaaacctttatigacaaagtattcaa 3103
Db 120 GGGAGATGGTAAAGAAACAATGAATGCTCTTTTTCAAACTTTATTGCACAAGTGATTTCAA 179
Qy 3104 gtcgtgttcaaaaaatatattcatgtacctgtatccagcagaagggaggttccagtc 3163
Db 180 GTCTGTGTTCAAAAATATATTATCATGTACTCTGATCCAGCAAGAGGGAGTCCAGTCAA 239
Qy 3164 gagtcactacaactgatttagttgttagaagaatgagaatggaacagtcaggaatggagg 3223
Db 240 GAGTCACACTACACTGATTTAGTGTGTTTAGAGAAATGAGAATGGAACAGTGAGGAATGGAGG 299
Qy 3224 ccattttccatgactctccctctgtaaacagaagaacagaagggagagagctggcctc 3283
Db 300 CCATATTTCCATGATCTCCCTTGTAAACAGAAGCAACAGAAGGGACAGAGCTGGCCTC 359
Qy 3284 tacatcactctcaccttccaaatctgtggaagtgcactctacttgcgaacacaaatcaa 3343
Db 360 TACATCACTCTCACCTTCCAAATCTTGTGGAAGTGCATCTACTTGCACAGAACCAAAATTA 419
Qy 3344 ctacttccaaagtctcgtgctctgcaggtggaactccagctcagctcgaaggaattaggaaa 3403
Db 420 CTTACTTCCAAAGTTCTGGCTGCTTGCAGGTGGAATCCAGCTTCCAGGAACAGAAATGGGTGAGGAA 479
Qy 3404 tgaaggtcttttttaaaagcttctcagcctctcctaggaacagaaatgggtgagccaa 3463
Db 480 TGAAGGTCTTTTTTAAAGCTTCTCAGCTTCTCAGGAAACAGAAATGGGTGAGCCTAA 539
Qy 3464 tctgcaatttctactacagcagcattgagaccagtttagattatgaaattattagagatt 3523
Db 540 TCTGCAATTTCTACTACAGGCATTTGAGACCAGTTAGATTTATTGAAATATTATAGAGATT 599
Qy 3524 atgaacacttaaatatgattgattgattgattgattgattgattgattgattgattgatt 3582
Db 600 ATGAACACTT-AATTATGATAGTGGTATGACATCGATGAGAACATGGGATACCTTTAGGAA 658
Qy 3583 gtgaattgacagggcata-ttagttgatgaatggagctcttttgagctctttaaagcc 3641
Db 659 GTAGAATTGACAGGCGCATATTAGTTGATGAATGGAGTCATTTGAGTCTCTTAATAGCC 718
Qy 3642 atgtatcataatt 3654
Db 719 ATGTCCTCACTAAT 731
RESULT 14
BG197194
LOCUS BG197194
DEFINITION RGT16433 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG197194
VERSION BG197194.1 GI:13718881
KEYWORDS EST.
SOURCE human.
```

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:27:55 ; Search time 5205.79 Seconds
(without alignments)
17699.435 Million cell updates/sec

Title: US-09-837-602-1

Perfect score: 4403

Sequence: 1 ttccgcacgagcgcggttg.....accgcgttgagctccagct 4403

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description

1	4204	95.5	4483	9	AF058696	AF058696	Homo sapi
2	4123.4	93.6	4412	9	AF051334	AB013139	Homo sapi
3	1862.8	42.3	56500	9	AF051339	AF069291	Homo sapi
C 4	1862.8	42.3	184919	9	AF069291	AC074178	Homo sapi
C 5	1861.2	42.3	107549	9	AC074178	AK001017	Homo sapi
6	1836.4	41.7	2044	9	AK001017	AF218575	Rattus no
7	1416.4	32.2	2605	10	AF218575	AF076687	Mus muscu
8	1400.4	31.8	2484	10	AF076687	AF092840	Mus muscu
9	1398.8	31.8	2473	10	AF092840	AB016988	Mus muscu
10	1395.6	31.7	2412	10	AB016988	AC023998	Homo sapi
11	692.8	15.7	64310	2	AC023998	AF230342	Gallus ga
12	684	15.5	2533	5	AF230342	AX192776	Sequence
C 13	552.2	12.5	646	6	AX192776	AX341490	Sequence
C 14	538.6	12.2	544	6	AX341490	AF049895	Homo sapi
C 15	451.2	10.2	330250	9	AF049895	AX067375	Sequence
16	358.4	8.1	396	6	AX067375	AC023998	Homo sapi
C 17	275	6.2	64310	2	AC023998	G66922	csnpnbsl-pc
C 18	273.4	6.2	434	11	G66922	G30107	human sts
C 19	264.4	6.0	421	11	G30107	G66921	csnpnbsl-pc
C 20	199.2	4.5	355	11	G66921	G66926	csnpnbsl-pc
C 21	163.4	3.7	296	11	G66926	AC004612	Homo sapi
C 22	163.4	3.7	88295	9	AC004612	G66925	csnpnbsl-pc
C 23	156.4	3.6	272	11	G66925	G66923	csnpnbsl-pc
C 24	136.4	3.1	325	11	G66923	AC110357	Rattus no
C 25	136	3.1	27768	2	AC110357	G66924	csnpnbsl-pc
C 26	134.4	3.1	244	11	G66924	G66920	csnpnbsl-pc
C 27	121.8	2.8	317	11	G66920	AC099905	csnpnbsl-pc
C 28	116.4	2.6	62686	2	AC099905	G66927	csnpnbsl-pc
C 29	103.6	2.4	293	11	G66927	AX345260	Sequence
30	97.6	2.2	6237	6	AX345260	AL031733	Human DNA
31	96.6	2.2	215861	9	HS4557	AL121773	Human chr
C 32	95.4	2.2	159267	9	CNS01DSE	AB038653	Homo sapi
C 33	95.4	2.2	175330	9	AB038653	AL133240	Human chr
34	95.4	2.2	196853	9	CNS01DUD	AX345261	Sequence
C 35	92.4	2.1	6237	6	AX345261	AP001881	Homo sapi
36	88.8	2.0	149100	9	AP001881	AP001282	Homo sapi
37	88.8	2.0	165258	9	AP001282	AC009363	Homo sapi
C 38	87.6	2.0	82707	9	AC009363	AL111167	Homo sapi
C 39	87.6	2.0	202267	9	DJ293M10	AL391294	Human DNA
C 40	87.2	2.0	152120	9	AL391294	AC102932	Mus muscu
41	87	2.0	34914	2	AC102932	AC027279	Homo sapi
C 42	82.8	1.9	194594	9	AC027279	AC055742	Homo sapi
43	82.6	1.9	181188	2	AC055742	AC007528	Homo sapi
44	82.6	1.9	188451	9	AC007528	AC092109	Homo sapi
C 45	82.6	1.9	204037	2	AC092109		

ALIGNMENTS

RESULT	1
AF058696	AF058696
LOCUS	Homo sapiens cell cycle regulatory protein p95 (NBS1) mRNA, 4483 bp
DEFINITION	complete cds.
ACCESSION	AF058696
VERSION	AF058696.1
KEYWORDS	GI:3098674
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 4483) Carney, J.P., Maser, R.S., Olivares, H., Davis, E.M., Le Beau, M., Yates, J.R. III, Hays, L., Morgan, W.F. and Petrini, J.H.
TITLE	The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response
JOURNAL	Cell 93 (3), 477-486 (1998)
MEDLINE	98250063
REFERENCE	2 (bases 1 to 4483)
AUTHORS	Carney, J.P., Maser, R.S., Olivares, H.A., Davis, E.M., Le Beau, M., Yates, J.R. III, Hays, L., Morgan, W.F. and Petrini, J.H.J.

[illegible]

Db 1561 ||||| TAAGGAGCAGCATCTATCTGAGAAATGAGCCTGTGGACACAAATCAGACAAATAACTATT 1620
QY 1621 tacagatacagatttaaaatctattgtgaaatatttcgcagtaaatctcattcctcgtcaga 1680
Db 1621 TACAGATACAGATTAAATCTATTGTGAAAAATCTCGCCAGTAAATCTCATCTCGAGA 1680
QY 1681 aaagctcaagatcaaaataaaaaagggaattggatgatgtggccatagaagatgaagtatt 1740
Db 1681 AAAGCTAAGATCAAAATAAAAAAGGGAATGGATGATGTGGCCATAGAAGATCAAGTATT 1740
QY 1741 ggaacsgttattcaaggacacaaaaaccagagttagaattgtagtgaagttcaaaaaca 1800
Db 1741 GGAACAGTTATTCAAGGACACAAAAACAGAGTTAGAAATGTGCTGAAAGTTCAAAAACA 1800
QY 1801 ggaaggaagatgtaattgttagaaaaaggccaagatggatataagaacaaatgacacttt 1860
Db 1801 GGAGGAAGATGTCAAATGTTAGAAAAAGGCCAAGGATGGATATAGAAACAAATGACACTTT 1860
QY 1861 cagtgatgaagcagtagccagaaagtagcaaaaatatctcaagaaaaatgaaattgggaagaa 1920
Db 1861 CAGTGATGAAGCAGTACCAGAAAGTAGCAAAATATCTCAAGAAAAATCAAAATGGGAAGAA 1920
QY 1921 acgtgaactcaagggaagactcaacta tggctcagctaaagaaatctatacaaatgacaaact 1980
Db 1921 ACGTGAACCTCAAGGAAGACTCACTATGTGTCAGCTAAAGAAATATCTAACAATGACAACT 1980
QY 1981 tcagagatgtagtgaagatcttccaaaaagctgttattgactgaatttagatcaactggt 2040
Db 1981 TCAGGATGATGATGTGAGATGCTTCCAAAAAGCTGTATTGACTGGAATTTAGATCACTGGT 2040
QY 2041 gattaaaaactctacttccagaaatccgtctggcatalaaatgtagtatgtgtcaactaaa 2100
Db 2041 GATTAAAACTCTACTTCCAGAAATCCGCTGCGCATAAATGATGATTGCGTCAACTAAA 2100
QY 2101 aaatttcaagaaattcaaaaaagctacatatccttgagcagggaacttccacacatcat 2160
Db 2101 AAATTTCAAGAAATTCAAAAGCTCACATATCCTTGAGCAGGAAAACTTCCACACATCAT 2160
QY 2161 tggagatcagatctaagctcatcgtctggaacaaatcagaaagatcacagaactagaagtggt 2220
Db 2161 TGGAGATCAGATCTAATAGCTCATCATGCTCGAAGAAATACAGAACTAGAAGATGGCT 2220
QY 2221 aagcgggaaatggaggtacaaatacaactgcaaaagaagctctctgtcgtgatct 2280
Db 2221 AAGGCAGGAAATGGAGGTACAAAATCAACATGCAAAAGAAAGACTCTTCTGATGATCT 2280
QY 2281 ttttagatadaatccctatttaaaaagggaagaaatgaatgaagattttaaagaagaacca 2340
Db 2281 TTTTAGATACAATCCTTATTATAAAAGGAGAAGATAACTGAGGATTTTAAAAAGAGCCCA 2340
QY 2341 tggaaaaactcctagtaagcatctacttcagcccaacagggttatataatataatagt 2400
Db 2341 TGGAAAAACTTCTAGTAAGCATCTACTTCAGGCCCAACAAGGTTATATGATATATAGTG 2400
QY 2401 tatagaagcgaatttaagttacaattgttttatggcctaataatttatataaataaaatgcacaa 2460
Db 2401 TATAGAAGCGATTAAAGTTACAAATGTTTATGCGCTAAATTTATTAATAAATATGCACAA 2460
QY 2461 aacttggattccttctgtatgtaacaattgttctgtcgttttcaggccttctgcatgcat 2520
Db 2461 AACTTTGATTCTTTGTATGTAACAATGTTTGTCTGTTTTCAGGCTTTGTCAATTGCAT 2520
QY 2521 cttttttcatttttaaaatgttttgtttattataaatagttataatagtcacagttcaaa 2580
Db 2521 CTTTTTTTCATTTTAAATGTGTTTGTGTTTATTAAATAGTTAATATAGTACACAGTTCAAA 2580
QY 2581 attctaaatracgtgaaggtaaaggactaaagtccaccttccaccattgtcctagctact 2640
Db 2581 ATTCTAAATGTACGTAAAGTAAA-GACTAAAGTCACCCCTTCCACCATTGTCCTAGCTACT 2639
QY 2641 ----- 2640

Db 2640 TGGTCCCTCCAGAAAAAAATTCATGACTACTCATTTCTTATGAATCTTTCCAGGATTTT 2699
QY 2641 -----tattttaataattctcacaaatgatagcataacata 2681
Db 2700 TGAGTCTCTATTCAAAATTCCTATTATTTTAAATTAATTTCTACAAATGATAGCATACAATA 2759
QY 2682 tgcagtggtctcacaccttgccttttacttagtaagaattaaaaattatagaatatacaat 2741
Db 2760 TGCAGTGTCTTACACCTTGCTTTTACTTAGT-AGATTAAAAATTAGAGAAATATCAAT 2818
QY 2742 ataagtgttttaataatttttttccattatgctgtagcttaccataaactctggtga 2801
Db 2819 ATAATGTTTTAAATATTTTCTTTTCCATTATGCTAGTCTTACCTAAACTCTGTGA 2878
QY 2802 tccaaacaaaatggtctcagtggtcagatgctcactcacatgttattctagctagtagaaa 2861
Db 2879 TCCAAACAAATGGCTTCAGTGGTGCAGATGTCACCTACATGTTATTTCTAGTACTAGAAA 2938
QY 2862 ctgaagaccatgtggagacttcatcaaacatgggttttagttttccaccagaattggaaagac 2921
Db 2939 CTGAAGACCATGTGGAGACTTCATCAAAACATGGGTTTAGTTTCCACCAGAATGGAAAGAC 2998
QY 2922 ctgtaccccttttgggtggtcttactgagctgggtgggtgctgttttggagcttatttag 2981
Db 2999 CTGTACCCCTTTTGGTGGTCTTACTGAGCTGGGTGGGTGCTGTTTGGACTTATTTAG 3058
QY 2982 agtctagtttctactataaagtagaaatggtgagattgtttcttttctacckta 3041
Db 3059 AGTCTTAGTTTCTTACTATTAAAGTAGAAATGGTGAATGTTTCTTTTCTTACCTTA 3118
QY 3042 aagggagatggtaagaacaaatgaatgctcttttttcaaacctttattgacaagtgttttc 3101
Db 3119 AAGGGAGATGGTAAGAAACAAATGAATGCTCTTTTTCAAACTTTATTGACAAGTCATTTTC 3178
QY 3102 aagtcgtgttcaaaaatatatactgtaactgtaactcagcagaagaaggaggttccagtc 3161
Db 3179 AAGTGTGTGTCAAAAATATATTCAATGACTGTGTGATCCAGCAGAAAGAGGAGTTCAGTC 3238
QY 3162 aagagtcactcacactgatttagttgttagaagtgaatggaacagtgaggaatgga 3221
Db 3239 AAGAGTCACCTACAACTGATTTAGTTGTTTAGAAGATGAGAATGGAACAGTGAGGAATGGA 3298
QY 3222 ggccatatttccatgacttcccttgttaacagaagcaacagaagggaagagctggcc 3281
Db 3299 GGCCATATTTCCATGACTTCCCTTGTAAACAGAAACACAGAAAGGACAAAGAGCTGGCC 3358
QY 3282 tctacatcactcaccacttccaaactgttggaagtgcatctacttccagagaaccacatt 3341
Db 3359 TCTACATCACTCTCACCTTCCAAATCTGTGGAAGTGCATCTACTTGCCAGAACCAAAAT 3418
QY 3342 aacttacttccagttctgctgaggtggaaactccagctgcaaggagtaggga 3401
Db 3419 AACTTACTTCCAGTTCTGGCTGCTGAGTGGAACCTCCAGCTGCAAGGAGTTAGGA 3478
QY 3402 aatgaagggtttttttaaagcttctcagccttcccttaggggaacagaaaaattgggtgagcc 3461
Db 3479 AATGAAGTCTTTTAAAAAGCTTCTCAGCCTTCTTAGGGAACAGAAAATTTGGGTGAGCC 3538
QY 3462 aacttgaaatttctactacaggcattgagaccagttagattattgaaatattatagagag 3521
Db 3539 AATCTGCAATTTCTACTACAGGCATTGAGACCAGTTAGATTATGAAATATTATAGAGAG 3598
QY 3522 ttatgaacacttaaatattatgtagtggtatgacattgtagaacaatgggatactttaga 3581
Db 3599 TTATGAACACTTAAATTTATGATAGTGTATGACATTTGGATAGAACATGGGATACTTTAGA 3658
QY 3582 agtagaattgacaggggcataattagttgataatggaggtcatttgagtcctttaaagcc 3641
Db 3659 AGTAGAATTGACAGGGCATATTAGTTGATGAATGGAGTCATTTGAGTCTCTTAAATAGCC 3718
QY 3642 atgtataataattaccaagtgaaagctgggtggaacatatggtctccattttacaggttaag 3701
Db 3719 ATGTATCATAAATTACCAAGTGAAGCTGGTGGAACATATGGTCTCCATTTTACAGTTAAGG 3778

Qy	3702	aataataatgacagattaataattgttctgtctatgcccacaatcccttttctaaggaaga	3761
Db	3779	AAATAATAGGACAGATTAAATATTTCTGTCTCATGCGCCACAATCCCTTTCTAAGGAAGA	3838
Qy	3762	ctgcctactactacagatttttatgttcaatttatgaataataagaatgagaggttc	3821
Db	3839	CTGCCCTACTATACAGTTTTTATTTGTCAATTTATGAATAATAAGATGAGAGATTTC	3898
Qy	3822	tggtaacctctgtctttacaaaattatgggtgtgtccagattttccctttttaaacmct	3881
Db	3899	TGGTACCTCCTGCTTTACAAATATGGGTGTTGCCAGTATTTTCCCTTTTAAACCAT	3958
Qy	3882	tcccaattcgggtgtgtagtgtaggtttccatttgggttttaattgtatatccctgat	3941
Db	3959	T-CCAATCGGTGTGTAGTGATG- - - - -TTTCATTTTGGTTTAAATTTGTATATCCCTGAT	4012
Qy	3942	agctataattgggtcatagaataattctttatatacatctatagatgaagctctctgcygata	4001
Db	4013	AGCTATAATTGGGTGCATAGAAAATCTTTATACATTTCTAGATGCAAGTCTCTTGTGCGATA	4072
Qy	4002	tacgtattgagatattacacactagctgtggttgactgtttcttttatgtcttttgatg	4061
Db	4073	TATGTATTGAGATATTACACCTAGTCTGTGGCTTGACTGTTTCTTTATGCTTTTGTATG	4132
Qy	4062	aatagaagttttaattttgacaaaggttcaaatattatttttttttttttttttttttttt	4121
Db	4133	AATAGAAGTTTAAATTTTGAAGAAGTCAAAATTTATTTTCTTTTCTTTTGTGATATTTT	4192
Qy	4122	tctctcaatttaaccacagatttcagatatctgtctctattatataaacctttatattt	4181
Db	4193	TCTCTCAAAATTAACCCCAAGATTTCAGATATTCTGTCTTATTATATAAACTTTATATT	4252
Qy	4182	ttatattgtgactacactgaattgatattgtattgttgaaattatgatacaggttctt	4241
Db	4253	TTATATTGATGATCTACTTGAATGATATGATGTTGTGAATTTATGATCAGGGTCTT	4312
Qy	4242	tttttccccatacaagatccagctattgttaacactgtttattgaaagaattatccctt	4301
Db	4313	TTTTTCCCCATACAAAGTATCCAGCTATTGTAACACTGTTTATTGAAAGAATTATCTCTT	4372
Qy	4302	cctcataaattacacttgcaatttagtaaaatacaataacacatrmammrrrgatc	4361
Db	4373	CCTCATTAATTTACCTTTGCCAAATTTAGTCTGTCGCCGA--ATTCCTGCAGCCCGGGGATC	4430
Qy	4362	cactagtctagagcgccgcaccgcggtggagctccagct	4403
Db	4431	CACGTAGTTCTAGAGCGCGCCGCCCGGGTGGAGCTCCAGCT	4472
RESULT 2			
AF051334			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			

JOURNAL	Submitted (26-FEB-1998) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr.11, Jena 07745, Germany									
FEATURES	Location/Qualifiers									
source	1. .4412									
gene	/organism="Homo sapiens"									
	/db_xref="taxon:9606"									
CDS	1. .4412									
	52. .2316									
	/gene="NBS"									
	/notes="mutated in Nijmegen Breakage syndrome"									
	/codon_start=1									
	/evidence=not_experimental									
	/product="nibrin"									
	/protein_id="AAC39732.1"									
	/db_xref="GI:3126795"									
	/translation="MWKLLPAAGPAGGEPYRLLTGVVVGRKNCAILIENDQISIRN HAULTANFSVTNLSQTDDEIPVFLKDKNSKYCTFVNEEKMQNGFSRLKSGDGIFFGVF GSFRIEYELVACSCDVSCKTALQALQUGGFTVNNTECTHLVMVSVKVTIK TICALICBPLRIVKPEYEFELKAVESKKQPPQIESFYPLDPEPSIGKSNVDSGRQER KQIFKGTFTFLNAKHKLSSAVVFGGEGEARLITEENEHNFFLAPGTCVDTGIT NSOTLIPDCOKWIOSIMDLORGLRPIPEAEIGLAVIEMTKNYCDPOGHPSTGLK TTTTPGSLSGSVDEKLMPSAPVNTTYYVADTSEQADTWLSEKPEIKVSKMEOK FRMLSQDAPTVKESCKTSSNNNSMVSNTLAKMRIPNTQLSPTKLPSPKSKDRASQOQ QTNISIRNYFQPTKKRERDEENQEMSSKSARIETSCSLLEQTQPTSPKSNKKEQHL SENEPVDNTSDNNLFTDILKSIVKNSASKSHAAEKLRSNKKREMDOVAIIDEVLEQL FKDTKPELEIDVRKQOEEDVNVKRRPMDIETNDTFSDEAVPESKISQENIEGKRR ELKEDSLWSAKEISNNDKLDDSEMLPKLLLTFRSLVTKNSTSRNPSGINDDYGOL KNFKPKKVTYPGAGKLPHIIGSDLLAHHARKNTELEEWLROEMEVONQHAKESLA DDLFRNPNYLRKR"									
	polyA_signal									
	2446. .2451									
	/gene="NBS"									
	/evidence=experimental									
	4392. .4397									
	/gene="NBS"									
	/evidence=experimental									
BASE COUNT	1470 a 753 c 853 g 1336 t									
ORIGIN										
	polyA_signal									
	polyA_signal									
	Query Match 93.68; Score 4123.4; DB 9; Length 4412;									
	Best Local Similarity 97.18; Pred. No. 0;									
	Matches 4287; Conservative 7; Mismatches 28; Indels 91; Gaps 6;									
QY	13	cgcggttgacgtcgcccgccagcctgagagcgccgacccgatgtggaactgctgccgc	72							
DB	12	CGCGGTTGCAGTCGCGCCCGAGCCCTGAGGAGCGCGGACCGATGTGGAACCTGCTGCCGC	71							
QY	73	cgcgggcccgcgagagagaaacacacagactttgactggcgttgagtagcgtttgttg	132							
DB	72	CGCGGGCCCGCAGGAGGAGAACCATACAGACTTTTGTACTGGCGTTGAGTACGTTGTTGG	131							
QY	133	aaggaaaaactgtgccattcttaattgaaatgatcagtcgatacgcgcaaatcatgctgt	192							
DB	132	AAGGAAAAACTGTGCCATTCTTGATTGAAAATGATCAGTCGATCAGCCGAAATCATGCTGT	191							
QY	193	gttaactgtctaaactttctgttaacacacactgagtcacacagatgaaatccctgtattgac	252							
DB	192	GTTAACTGCTAACTTTTCTGTAAACCACTGAGTCAACACATGAATCCCTGTATTGAC	251							
QY	253	attaaagaataattctaagtatggtacactttgttaatgaggaataatgcagaatggcct	312							
DB	252	ATTAAGAATAATTCTAAGTATGTTACCTTTGTTAATGAGGAAAAATGCAGAAATGGCTT	311							
QY	313	ttccggaactttgaagtcgggggagtggtattacttttggagtggtttggaagtaaatcag	372							
DB	312	TTCCCGAACCTTTGAAGTCGGGGGATGTTACTTTTGGAGTGTGTTGGAAGTAAATTCAG	371							
QY	373	aatagatgatgcctttgttgatgctctctcttcttcttcttcttcttcttcttcttctc	432							
DB	372	AATAGATGATGAGCCCTTGGTTGTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC	431							
QY	433	tttaaatcaagctatatgtgcaacttggaggatttactgtataacaattggacagaagaatg	492							

Db 432 TTTAAATCAAGTATATATTGCAACTTGGAGGATTTACTGTAAACAATTTGACACAGAAGATG 491
Qy 493 cactcacccttgcatgggtatcagtgaaagttaaccatttaaaacaataatgtaactcaatttg 552
Db 492 CACTCACCCTTGTCATGGTATCATAGTGAAGGTTACCATTAAACAATAATGTGCACTCAATTTG 551
Qy 553 tggagctccaattgtaagccagaataattttactgaattccctgaaagcaggttcagtcocaa 612
Db 552 TGGACSTCCAATTGTAAGCCAGAGATATTTTACTGAATTCCTGAAAGCAGTTGAGTCCAA 611
Qy 613 gaagcagcctccaataattgaaagtttttaccacaacctcttgatgaaccatctattggaag 672
Db 612 GAAGCAGCCTCCACAAATTTGAAAGTTTATACCACCTCTTGATGAACCATCTATTGGAG 671
Qy 673 taaaaatgtgtctgtcagcgccaggaagaaagaaacaaactctcaagggaaacatt 732
Db 672 TAAAATGTTGATCTGTCAGGACGGCAGGAAAGAAACAAATCTTCAAAGGAAACAAT 731
Qy 733 tatattttgaatgccaaacagcataaagaatttgagttccgcagttgtcttggagtg 792
Db 732 TATATTTTGAATGCCAAACAGCATAGAATAATTGAGTTCGCGCAGTTGTCTTTGGAGGTGG 791
Qy 793 ggaactaggttgataacagaagagaatgaagaagacaataatttcttttggctccgg 852
Db 792 GGAAGCTAGGTTGATAACAGAAGAGAATGAAGAACAATAATTTCTTTTGGCTCCGG 851
Qy 853 aacgtgtgtgttgatacaggaataacaaactcacagaccttaattctctgactgtcagaa 912
Db 852 AACGTGTGTTGTATACAGGAATTAACAACATCACAGACCTTAATTTCTTGACTGTGAGAA 911
Qy 913 gaaatggattcagtcataatggatatgctccaaaggcagggtctcttagacctattctga 972
Db 912 GAAATGGATTCAATTAATGGATATGCTCCAAAGGCCAAGGTCITTAGACCTATTTCCTGA 971
Qy 973 aqcaaaaatttgattggcggtgatttctagctacaagaatactactgtgactcctcaggg 1032
Db 972 ACAGAAATTTGGAATTTGGCGGTGATTTTCATGACTACAAAGAAATTAAGTGTGCTCCAGGG 1031
Qy 1033 ccattccagtcagatgaagacaacactccaggaccaaagccttccaaagggcgtgctc 1092
Db 1032 CCATCCAGTACAGGATTAAGACACAACTCCAGGACCAAGCCTTTTCAAGGCGGTGTC 1091
Qy 1093 agttgatgaaactaattgccagcgcccgagtgacactacaatacgtactgtgacac 1152
Db 1092 AGTTGATGAAAACTAAATGCCAAGGCCCCAGTGAACACTACAACATACGTAGCTGACAC 1151
Qy 1153 agaatcagaagcagagatacatgggatttgagtgaaagccaaagaaatacaagctctc 1212
Db 1152 AGAATCAGAGCAAGCAGATACATGGGATTTGAGTGAAGGCCAAAGAAATCAAAAGTCTC 1211
Qy 1213 caaaatggaacaaaaattcagaatgctttcaagaagcagccactgtaaaagagtcctg 1272
Db 1212 CAAAATGGAAACAAAATTCAGAAATGCTTTTCAAGAGTGCACCCACTGTAAAGGAGTCCGTG 1271
Qy 1273 caaaacagctctaataataatagtgatgatacaatactttggctaaagttagaataccc 1332
Db 1272 CAAAACAAGCTCTAATAATAATAGTATGATATCAAAATCTTTGGCTTAAGATGAGAATCCC 1331
Qy 1333 aaactatcagctttccaccaactaaattggccaagataaaataaagttaagtagggcttc 1392
Db 1332 AAACATATCAGCTTTTACCACAACTAAATTTGCCAAGTATATAATAAGGTAAAGATAGGGCTTC 1391
Qy 1393 tcagcagcagcagacaactccatcagaactactcttcagcgctctaccaaaaaaggga 1452
Db 1392 TCAGCAGCAGCAGACCACTCCATCAGAACTACTTTTACGCCGTCTACCAAAAAAGGGA 1451
Qy 1453 aagggatgaagaaaatcaagaaatgcttctcatgcaaatcagcaagataagaacgctcttg 1512
Db 1452 AAGGGATGAAGAAATCAAGAAATGCTTTCATGCAAAATCAGCAAGAAATAGAACGCTTTG 1511
Qy 1513 ttctcttttagaacaacaacacctgctacacccctcattgtgaaaaataaaggagcagca 1572
|||||

Db 1512 TTCTCTTTTAGAAAAACACAACTGCTACACCCCTCATTTGTGGAAAAAATAAGGAGCAGCA 1571
Qy 1573 tctatctgaaatgacccctgtgacacaaactcagacaaataacttatttacacatacaga 1632
Db 1572 TCTATCTGAAATGACCTGTGACACAAACTCAGACAAATAACTTATTTACAGATACAGA 1631
Qy 1633 tttaaaaatctattgtgaaaaattctccagtgaaatctctactgctgcagaaaaagctaaagtc 1692
Db 1632 TTTAAAAATCTATTGTCAAAAATTTCTGCCAGTAAATCTCATGCTGCAGAAAAAGCTAAGATC 1691
Qy 1693 aaataaaaaaggggaaatggatgagtggccatagaagatgaagtatttgggaaacagttatt 1752
Db 1692 AAAATAAAAAAGGAAATGGATGATGTGCCATAGAAGATGAAGTATTGGAAACAGTTATT 1751
Qy 1753 caaggcacaaaaccagagttagaatttgatgtgaagtttcaaaaacagagagaagatgt 1812
Db 1752 CAAGGACACAAAACAGAGTTAGAAATTTGATGTGAAAGTTCAAAAACAGGAGGAAGATGT 1811
Qy 1813 caatgttagaaaaagggccaaggatggatatagaaaacaaatgcacacttccagtgatgaagc 1872
Db 1812 CAATGTTAGAAAAAGGCCAAGGATGGATATAGAAACAATAATGCACACTTTCAGTGATGAAGC 1871
Qy 1873 agtacccagaaagttagcaaaatatctcaagaaaaatgaaattgggaagaaaacgtgaactcaa 1932
Db 1872 AGTACCAGAAAGTAGCAAAATATCTCAAGAAAATGAAATTTGGGAAGAAACGTGAACCTCAA 1931
Qy 1933 ggaagactcactatggtcagctaaagaaatatctaaacaatgacaactccagactcagatgag 1992
Db 1932 GGAAGACTCATTATGTGTGCTAGCTAAGAAATATCTACAAATGACAAACTTTCAGGATGATAG 1991
Qy 1993 tgagatgcttccaaaaagctgttattgactgaaatttagactcactcgtggtgattaaaaactc 2052
Db 1992 TGAGATGCTTCCAAAAAGGCTTATTGACTGAATTTAGATCCTACTGGTGATTAATAAACTC 2051
Qy 2053 tacttccagaataatccgtctggcataaataatgatgatattatggttgcactaaaaatttcaagaa 2112
Db 2052 TACTTCCAGAAATTCATCTGGCATAAATGATGATTAATGTGTCAACTAAAAATTTTCAAGAA 2111
Qy 2113 attcaaaaagggtcacatatctctggagcaggaagaaactccacacatcatttgagagatcaga 2172
Db 2112 ATTCAAAAGGTCACATATCTCTGGAGCAGGAAACCTTCCACACATCATTTGGAGGATCAGA 2171
Qy 2173 tctaaatagctcatcactgctgaaagaataacagaactagaaagatggtggttaagcaggaaat 2232
Db 2172 TCTAATAGCTCATCATGCTCGAAGAAATACAGAACTAGAAAGTGGCTAAGCGCAGGAAAT 2231
Qy 2233 ggaggtacaaaatcaacatgcaaaagaagatctcttctgctgatatcttttttagatacaa 2292
Db 2232 GGAGGTACAAAATCAACATGCAAAAAGAGAGTCTCTTGGCTGATGATCTTTTTTAGATACAA 2291
Qy 2293 tccttatttaaaaaggagaagataaactgaggtatttlaaaagaagccatggaaaaaactc 2352
Db 2292 TCCTTATTTAAAAGGAGAGATAACTGAGGATTTTAAAAGGAAGCCATCGAAAAAACTTC 2351
Qy 2353 ctagtgaagcatctacttcaggcccaacaaggttataagaatatagtgatagagaagocat 2412
Db 2352 CTAGTAAGCATCTACTTTCAGGCCAACAAAGGTTATATGAATATATAGTGTATAGAAGCGAT 2411
Qy 2413 ttaagttacaagtcttttatggcctaataatttattaaataaaatgcacaaaactttgattct 2472
Db 2412 TTAAGTTACAATGTTTTTATGGCCCTAAATTTATTAATAAATGCACAAAATTTTGATPCT 2471
Qy 2473 ttgtgtgtaacaattgtttgtctgttttcaggctttgtcattgcatcttttttctatt 2532
Db 2472 TTTGTATGTAACAATTTGTTTGTCTGTTTTCAGGCTTTGTTCATTGTCATCTTTTTTTCATT 2531
Qy 2533 tttaaatgtgttttgtttatttaaatagtttaatatagtcacagtttcaaaatctcaaatrta 2592
Db 2532 TTTTAATGTGTTTGTTTTATTAATAATAGTTAAATAGTACAGTTCACAAATTTCTAAATGTA 2591
Qy 2593 cgttaaggtaaggaactaaagtcaccttccaccttgcaccttgcctactact----- 2640
Db 2592 CGTAAGGTAAA-GACTAAAAGTCACCTTCCACCATTGTCCTAGCTACTTGGTTCCTCCTCA 2650
|||||

```
Qy 2641 ----- 2640
Db 2651 GAAAAAATTCATGACTACTCAATTTCTTATGAATCTTCCAGGGATTTTGTAGTCCATATPC 2710
Qy 2641 -----tattttfaataatttctcacacaaatgatagcataacatcagtggttcta 2693
Db 2711 AAATTCCTATATTTTAAATAATTTCTACACAAATGATAGCATAACATATGAGTGTTCTA 2770
Qy 2694 caccttgctttttacttagtaagataaataattataggaataatcaataatgttttta 2753
Db 2771 CACCTTGCTTTTACTTAGT - AGATTAAAAAATTATAGGAATAATCAATATAATGTTTTA 2829
Qy 2754 atatttttttccattatgctgtagtcttaactaaactcctggtgatccaaacaaat 2813
Db 2830 ATATTTTCTTTTCCATTAATGCTGATGCTTACTTAACCTCTGGTGATCCAAACAAAT 2889
Qy 2814 ggtctcagtggtgcagatgtcacctcacatgatttcttagtactagaacactgaagaccatg 2873
Db 2890 GGCTTCAGTGGTGAGATGTCACCTACATGTTATTTCTAGTACTAGAAACGAAAGACCATG 2949
Qy 2874 tggagacttcatcaaacatgggttttagtttccacagaaatggaaagacctgtaccccttt 2933
Db 2950 TGGAGACTTCATCAACATGGGTTTAGTTTTTACCAGAAATGGAAGACCTGTACCCCTTT 3009
Qy 2934 ttggtgcttactagactgggtggtgctgtgtttgagcttattagagtcctagtttt 2993
Db 3010 TTGGTGCTCTTACGAGCTGGGTGGGTGCTGTTTTGAGCTTAATTTAGAGTCCCTAGTTTTT 3069
Qy 2994 cctacttataaagttagaaatggtgagatggtttctttctttctcactcktaaaaggagatggt 3053
Db 3070 CCTACTTATAAGTAGAAATGGTGAGATGTTTCTTTTCTACTACCTTAAGGGAGATGGT 3129
Qy 3054 aagaacaaatgaatgcttttttccaaacttattgacaagtatttccaagtctgtgttc 3113
Db 3130 AAGAAACAATGAATGTCTTTTTCAAACTTTAATTGACAAGTGATTTTCAAGTCTGTGTTCT 3189
Qy 3114 aaaaatatattcatgtacctgtcatccagcaagaaggaggttccagttcaagagtcactac 3173
Db 3190 AAAAATATATTCATGTACCTGTGATCCAGCAAGAAGGGAGTTCAGTCAAGAGTCACTAC 3249
Qy 3174 aactgattagtcttagagaatgagaaatggaaactggaacagtgaggaatggagcccatattcc 3233
Db 3250 AACTGATTAGTGTGTTAGAAATGAGAAATGGAACAGTGAAGGAATGAGGCCATATTTCC 3309
Qy 3234 atgaactccctgtaaacagaagaacagagaagggacaagaggtgacctctacatcactc 3293
Db 3310 ATGACTTCCCTTTGTAACAGAAACAACAGAAAGGCAAGAGGCTGGCCTCTACATCACTC 3369
Qy 3294 tcaccttccaaatcttgtggaagtgcactcacttgcagaaaccaaaatctaacttactcca 3353
Db 3370 TCACCTTCCAAATCTTGTGGAAGTGATCTTACTTGCCAGAACCAAAATTAACCTTACTTCCA 3429
Qy 3354 agttctggctgcttcaggttggaaactccagctcgaaggaggttagggaaatgaagtcctt 3413
Db 3430 AGTTCTGGCTGCTTGAGGTGGAACCTCCAGCTCGAAGGGAGTTAGGGAATGAAGTCTT 3489
Qy 3414 tttttaaagctctcagccttccctaggaagaacagaataatgggtgagccaatctgcaattt 3473
Db 3490 TTTTAAAAAGCTTCTCAGGCTTCTAGGGAACAGAAATTTGGGTGAGCCAATCTGCAATTT 3549
Qy 3474 ctactacagcattgagaccagttagattattgaaatattatagagagttatgaacactt 3533
Db 3550 CTACTACAGCATTTGAGACCGATTAGATTATTGAAATATATTATAGAGAGTTATGAACACTT 3609
Qy 3534 aaattatgtagtgggtgacattggatagaacatgggataccttttagaagtagaattgac 3593
Db 3610 AAATTATGATAGTGTATGACATTGGATGAACATGGGATCTTTTAGAAGTGAATTGAC 3669
Qy 3594 agggcatattagttgataagaaaggagtcatttgagtcctytaaataagccatgtatcataat 3653
Db 3670 AGGGCATATTAGTTGATGAATGGAGTCAATTTGAGTCTCTTAAATAGCCATGTATCATAAT 3729
```

```
Qy 3654 taccaaagtaagctggttggaacatatggtctccatttttacaggttaagaataataatggac 3713
Db 3730 TACCAAGTGAAGCTGGTGGAAACATATGCTCTCCATTTTACAGTTTAAGGAATAATAATGGAC 3789
Qy 3714 agattaatatgttctgtctcatgccacaatccctttcttaaggaagactgcccactactat 3773
Db 3790 AGATTAAATTTGTTCTCTGTCATGCCCAATCCCTTTCTTAGGAAGACTGCCCTTACTAT 3849
Qy 3774 agcagtttttattattgttcaatttatgaataataatgaatgagggagttctctggtacccctg 3833
Db 3850 AGCAGTTTATTAATTTGTCATTTATGAATTAATGAATGAATGA - GAGTTCTGGTACCTCCTG 3908
Qy 3834 tctttacaaaatbattgggtggtccagbatttttcccttttttaacccmttcccaatccggg 3893
Db 3909 TCTTTACAAATATGG-----TGTGTCAGTATTTTTCCTTTTAAACCATTCCAATCGGTG 3964
Qy 3894 tgtgtagtggtgattttccatttgggttttaatttggatatccctcgatagctataatggg 3953
Db 3965 TGTAGTGATG-----TTTCATTTTGGGTTTAAATTTGTATATATCCCTGTAGTATATAATTGG 4019
Qy 3954 gtcatagaaaatctttatcacattctagatgcaagctctctgycggatatacagattgaga 4013
Db 4020 GTCATAGAAATCTTTATACATTTCTAGATGCAAGTCTCTTGTCCGATATATGATGAGA 4079
Qy 4014 tattacacctagctgctggttgactggtttcttttatgtcttttttgatagaatagaagtttt 4073
Db 4080 TATTACACCTAGTCTGTGGCTTGACTGTTTCTTTTATGCTCTTTTGATGAATGAAGAAGTTT 4139
Qy 4074 aaattttgacaaggtcaaaatttatttttttttttttttttttttttttttttttttttt 4133
Db 4140 AAATTTTGACAAGGTCAAAATTTATTTTCTTTTGTGTGATATTTTCTCTCCAAATTT 4199
Qy 4134 aaccccaagatttcagatatattctgtctctattatatataaaactttattttttattttgtga 4193
Db 4200 AACCCCAAGATTTCCAGATATTTCTGCTCTATATATATAAATTTATATTTATATTTGGA 4259
Qy 4194 tctacctggaatgatatgtagtctgtagaattatcgatcagatcaggggttcttttttccccc 4253
Db 4260 TCTACCTTGAATTGATATGATGTGTGAATTTATGGATCAGGGTCTCTTTTTCCTCCCAT 4319
Qy 4254 acaagttacagtcattgtaacactggtttattgaaagaattatccttctcccatataatt 4313
Db 4320 ACAAGTATCCAGTCATTTGTAACACTGTTTATTTGAAAGAAATATCCTTTCCTCAATAAT 4379
Qy 4314 accttgccaattagtaaaaaatcaaatcaaccat 4346
Db 4380 ACCTTGCCAATTAGTAAAAAATCAATTAACCAT 4412
RESULT 3
AB013139
LOCUS
DEFINITION Homo sapiens gene for NBS1, complete cds.
ACCESSION AB013139
VERSION AB013139.1 GI:3169124
KEYWORDS NBS1; Nijmegen breakage syndrome.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
Matsuura,S., Tauchi,H., Nakamura,A., Kondo,N., Sakamoto,S.,
Endo,S., Sneets,D., Solder,B., Belohradsky,B.H., Kaloustian,V.M.,
Oshimura,M., Isomura,M., Nakamura,Y. and Komatsu,K.
Positional cloning of the gene for Nijmegen breakage syndrome
Nat. Genet. 19 (2), 179-181 (1998)
98282099
REFERENCE
2 (bases 1 to 56500)
Matsuura,S., Tauchi,H. and Komatsu,K.
Direct Submission
Submitted (21-APR-1998) Shinya Matsuura, Hiroshima University,
Research Institute for Radiation Biology and Medicine, Department
of Radiation Biology; Kasumi 1-2-3, Minami-Ku, Hiroshima, Hiroshima
```



```
/db_xref="GI:3687831"
/translation="MKLPARVFTLGSRLPCGLAPRRFFSYGKILYQNFEALQSRFF
SPIQKAMLPNSPQGVAFITGGTGLGKGMTLLSLGAQCQVIASRKMDVLKATAEQ
ISSQTGNKHVIAIOCDVRDPMQVNTVELIKVAGHPNIVINNAAGNFISPTERLSPNA
WKITIDVLHNGTAFVTLIEIGKQLIKAQK"
176204..176406
/exon
/gene="DECR"
/number=2
177989..178046
/exon
/gene="DECR"
/number=3
178166..178252
/exon
/gene="DECR"
/number=4

Query Match      42.3%; Score 1862.8; DB 9; Length 184919;
Best Local Similarity 94.1%; Pred. No. 0;
Matches 2024; Conservative 9; Mismatches 26; Indels 91; Gaps 6;

QY 2278 tcttttagatacaatccttatttaaagaagagaagataactgaggatttttaaaagaag 2337
DB 94701 TCTTTGCAGATACAACTCTTATTAAAGAGGAGAAGATAACTGAGGATTTTAAAAAGAG 94642
QY 2338 ccattgaaaacttcctagtaagcatctacttcaggccacaaggttatatgaatatata 2397
DB 94641 CCATGGAAAAACCTTCCTAGTAGAAGCATCTACTTCAGGCCAACAGGTTATATGAATATATA 94582
QY 2398 gtgtatagaagcgattaaagttaacaattttttatggcctaatttataataaaatgca 2457
DB 94581 GTGTATAGAAGCGATTAAAGTTACAAGTTTATGGCCTAAATTTATTAATAAAGTCA 94522
QY 2458 caaaacttgattctttgtatgatacaaatgtttgttyctgttttcaggcctttgtaattg 2517
DB 94521 CAAAACCTTTGATCTTTTGTATGATGAACAAATTTGTTTGTCTGTTTTCAGGCTTTGTCATTG 94462
QY 2518 catcttttttcatttaaatgtttttgttttatttaaatagtttaataatgaatcagttc 2577
DB 94461 CATCTTTTTCATTTTAAATGTTGTTTGTATTAAATAGTTAAATATAGTCACAGTTC 94402
QY 2578 aaaattctaarrtcacgttaaggtaaaagactaaagtacacccttcaccattgtctcagct 2637
DB 94401 AAAATTCTAAATGACGTAAAGTAAA -GACTAAAGTCACCCCTCCACCATTGCTCCTAGCT 94343
QY 2638 act----- 2640
DB 94342 ACTTGGTTCOCCTCAGAAAAAAATTCATGACTACTTCTTATGAATCTTTCAGGCGAT 94283
QY 2641 -----tattttaaatatttcctacacaataatgatacataac 2678
DB 94282 TTTTGAGTCCCTATTCAAAATTCCTATTTTTAAATTAATTTCTTACACAATGATAGCATAAC 94223
QY 2679 atatgcagtggttlacacacctgttttttacttagtaagatttaaaattataggaatatc 2738
DB 94222 ATATGCAGTGTCTACACCTGCTTTTACTTAGT -AGATTAAATATATAGGAATATC 94164
QY 2739 aataaatgttttbaatatattttctttccattatgtctgtagcttaccetaaacctcgg 2798
DB 94163 AATAAATGTTTTTAAATATTTTCTTTTCCATTATGCTGTAGTCTTACCTFAAACTCTGG 94104
QY 2799 tgatccaaaacaaatggcttcaggtgagatgctacacctacatggttatcttagtactag 2858
DB 94103 TGATCCAAAACAAAATGGCTTCAGTGGTGCAGATGTCACCTACATGTTATTCTAGTACTAG 94044
QY 2859 aaactgaagaccatgtgagacttcatacaacatgggttttagttttccaccagaatggaaa 2918
DB 94043 AAACCTGAAGACCATGTGGAGACTTCATCAACACATGGGTTTATAGTTTTCACCAGAAATGAAA 93984
QY 2919 gacctgaccccccttttgggtgttctactgagctgggtgggtgctgttttggagcttatt 2978
DB 93983 GACCTGTACCCCTTTTGGTGGTCTTACTGAGCTGGGTGGGTGCTGTGTTTGAGCTTATT 93924
QY 2979 tagagctcctagtttctactataaagtagaagaatggtgagattgttttctttcttacc 3038
|||||
```

```
DB 93923 TAGAGTCTCTAGTTTTCCTACTATTATAAAGTAGAAAATGGTGAGATGCTTTTCTTTTCTTACC 93864
QY 3039 ktaaggagagtggttaagaacaacaatgaatgtcttttttccaactttattgacaagtgtt 3098
|||||
DB 93863 TTAAAGGGAGATGTTAAGAAACAATGAATGCTTTTTCAAACTTTATTGACAAGTGATT 93804
QY 3099 ttcaagtctgttgcataaaaaatatattcatgtacctgtgtatccagacaagaagggagttcca 3158
|||||
DB 93803 TTCAAGTCTGTGTTCAAAAATATATTTCATGTACCTGTGTATCCAGCAGCAAGAGGAGTTCCA 93744
QY 3159 gtcaagagtcactcaactgattagttgttttagagaaatgagaaatggacaagtggaggaat 3218
|||||
DB 93743 GTCAAGAGTCACTACACTGATTAGTTGTTTAGAGAAATGAGAAATGGAACTAGTAGGAAT 93684
QY 3219 ggaggccatatattccatgaactcccttgaacagaagaacaacagaggacaagagctg 3278
DB 93683 GGAGGCCATATTTCATGACTTCCTCTGTAAACAGAAACAACAGAGGACAGAGGCTG 93624
QY 3279 gcctctacatcactctcactctccaaatccttgtggaagtcatctactctgcagaaaccaa 3338
|||||
DB 93623 GCCTCTACATCACCTCTCACCTTCCAAATCTTGTGGAAGTGCATCTACTTGCAGAACCAA 93564
QY 3339 attaaacttactccaagttctgctgcttcaggtggaactccagctcgaagggagttag 3398
|||||
DB 93563 ATTAACTTACTTCCAAAGTTCTGCTGTCAGAGTGGAACTCCAGCTCGAAGGGAGTTAG 93504
QY 3399 ggaatgaaggtcttttttaaaagcttcagcgccttcctcagggaacagaattgggtga 3458
DB 93503 GGAAATGAAGTCTTTTAAAAAGCTTCTCAGCCTTCTAGGGAACAAGAAATTTGGGTGA 93444
QY 3459 gccaatctgcgaatttctactcacagggcattgagaccagttagattattgaaattattataga 3518
DB 93443 GCCAATCTGCAATTCTACTACAGCATTGAGCCAGTTAGATTATTGAAATATTATATAGA 93384
QY 3519 gagttatgaacacactaaattatgatagtggtatgacattgataagacaatgggacacttt 3578
DB 93383 GAGTTATGAACACCTTAAATTTATGATAGTGGTATGACATTTGGATAGAACATGGGATACATT 93324
QY 3579 agaagtagaattgacagggcataattagttgataaaggagtcattttagtctctytaaata 3638
DB 93323 AGAAGTAGAATTGACAGGGCATATTAGTTGATGAAATGGAGTCAATTTTCAGTCTCTTAATA 93264
QY 3639 gccatgtatcataattaccagaagtgaagctggtggaacaatatgggtctcattttacagttta 3698
DB 93263 GCCATGTATCATAAATTACCAAGTGAAGCTGGTGGAAACATATGGTCTCCATTTTACAGTTA 93204
QY 3699 aggaataatggacagattaaattgttctgctgcacgccacaacatcccttctcaagga 3758
DB 93203 AGGAATATAATGGACAGATTAAATATTGTTCTCTGTCATGCCCAACANTCCCTTCTTAAGA 93144
QY 3759 agactgcctactatagcagtttttataattgtcaatttatgaataataatgaatgagagag 3818
DB 93143 AGACTGCCCTACTATAGCAGTTTATATATTGTCAAATTTATGAATATAATGAATGA -GAG 93085
QY 3819 ttctgggtacctcctgtcttttacaataattgggtgtgtgtccagtagtatttttcccttttaac 3878
DB 93084 TTTCTGGTACCTCTCTCTTTACAAAATATTGG -TGTGTGTCAGTATTATTTTCTTTTAA 93029
QY 3879 omttcccaattccgggtgtgtaggtgagtttccatttgggtttttaaatttatataccct 3938
DB 93028 CCATTCCAACCGGTGTAGTGATG -TTTCATTTTGGTTTAAATTTGTTATATCCCT 92974
QY 3939 gatagctataattgggttcataagaaattctttatacattcttagatgcaagctccttgcgg 3998
DB 92973 GATAGCTATAATTGGGTTCATAGAAATCTTTTATACATTTAGATGCAAGTCTCTGTGTCGG 92914
QY 3999 atatacgtattgagatatcacacctagctgtgtggtgactgttttcttcttbtgtttt 4058
DB 92913 ATATATGTTATGAGATATTACACCTAGTCTGTGGCTTGACTGTCTTTTATGTCTTTT 92854
QY 4059 atgaatagaagttttaaattttgacaaggtcaaaatttttttttttttttttttttttttttt 4118
DB 92853 ATGAATAGAAGTTTTTAAATTTTGACAAGGTCAAAATTTTATTTTCTTTTGTGTTGATATT 92794
```



```

repeat_region      /rpt_family="Alu"
7558..7871
repeat_region      /rpt_family="Alu"
7938..8238
repeat_region      /rpt_family="Alu"
9286..9467
repeat_region      /rpt_family="MIR"
9941..10012
repeat_region      /rpt_family="MIR"
11013..11334
repeat_region      /rpt_family="Alu"
11651..11957
repeat_region      /rpt_family="Alu"
12133..12441
repeat_region      /rpt_family="Alu"
12513..12804
repeat_region      /rpt_family="Alu"
12808..12912
repeat_region      /rpt_family="L1"
12913..13213
repeat_region      /rpt_family="Alu"
13214..13638
repeat_region      /rpt_family="L1"
13639..13750
repeat_region      /rpt_family="L2"
14231..14529
repeat_region      /rpt_family="Alu"
16063..16506
repeat_region      /rpt_family="L1"
17682..17990
repeat_region      /rpt_family="Alu"
18183..18704
repeat_region      /rpt_family="ERV"
18757..18921
repeat_region      /rpt_family="MIR"
18939..19063
repeat_region      /rpt_family="Alu"
19103..19178
repeat_region      /rpt_family="MIR"
19312..19448
repeat_region      /rpt_family="MIR"
19626..19961
repeat_region      /rpt_family="L2"
20170..20245
repeat_region      /rpt_family="MER1_type"
20787..20974
repeat_region      /rpt_family="L1"
21556..21787
repeat_region      /rpt_family="L1"
21793..21917
repeat_region      /rpt_family="Alu"
21918..22210
repeat_region      /rpt_family="Alu"
22211..22380
repeat_region      /rpt_family="Alu"
22401..22804
repeat_region      /rpt_family="L1"
22971..23110
repeat_region      /rpt_family="L2"
23617..23881
repeat_region      /rpt_family="Alu"
24264..24633
repeat_region      /rpt_family="MalR"
24697..24908
repeat_region      /rpt_family="MIR"
25116..25179
repeat_region      /rpt_family="L1"
28103..29649
repeat_region      /rpt_family="L1"
29677..31685
repeat_region      /rpt_family="L1"
31689..31818
repeat_region      /rpt_family="Alu"

```

```

repeat_region      31819..32119
Query Match       42.3%; Score 1861.2; DB 9; Length 107549;
Best Local Similarity 94.1%; Pred. No. 0;
Matches 2023; Conservative 9; Mismatches 27; Indels 91; Gaps 6;

Qy 2278 tcttttagatacaatccttatttaaaagagagaagataaactagaggatttttaaaagaag 2337
Db 68939 TCTTTGCAGATACAACTCTTATTTAAAAAGGAGAAGATAACTAGAGATTTTAAAAAGAAG 68880

Qy 2338 ccattgaaactccctagtaagcatctacttcaggccaacagggtatatgaatatata 2397
Db 68879 CCATGGAAAAAATCTCCTAGTAGCATCTACTTCAGGCCAACAAAGGTTATATGAATATATA 68820

Qy 2398 gtgtagaagcgaattaaagttaacaatttttataggcctaaattttataaataaatgca 2457
Db 68819 GTGTATAGAAGCGATTAAAGTTACAATGTTTATGGCCCTAAATTTATTAATAATAATGCA 68760

Qy 2458 caaaactttgattcttttgatgatacaaatgttttgyctgttttcaggctttgcatgg 2517
Db 68759 CAAAACCTTTGATTCTTTTGTATGTAACAATTTGTTTGTCTGTTTTCAGGCTTTGTCTATTG 68700

Qy 2518 catcttttttcatttttaaatgtgtttgttttataaataagtttaataagtcacagttc 2577
Db 68699 CATCTTTTTCATTTTAAATGTGTTTGTATTATTAATAAGTTAATAATAGTCACAGTTC 68640

Qy 2578 aaaattctaaatrtaagtaagtaaaagtaaaagtaaaagtaaaagtaaaagtaaaagtaaa 2637
Db 68639 AAAATCTTAATATACGTAAAGTAAA-GACTAAAGTACCCCTTCACCATTTGTCTTAGCT 68581

Qy 2638 act----- 2640
Db 68580 ACTTGGTTCCCTCAGAAAAAATTCATGATACATCTTCTTATGAATCTTTCAGGGAT 68521

Qy 2641 -----tattttaataatttcctacacaaatgatacataac 2678
Db 68520 TTTTGAGTCTCTATTCAAATTCCTATTATTTTAAATAATTTCTTACACAAATGATAGCATAAC 68461

Qy 2679 atatgcagtgctcacacctgtcttttacttagtaagattaaaaattataggaataac 2738
Db 68460 ATATGCAGTGTCTTACACCTTGCTTTTACTTAGT-AGATTAAAAATTTATAGGAATATC 68402

Qy 2739 aataaatgtttttaatatatttttttccattatgtctgtagcttaccctaaactctgg 2798
Db 68401 AATAATAATGTTTAAATAATTTTCTTTTCCATTATGCTGACTCTTACCTTAAACTCTGG 68342

Qy 2799 tgatccaaacaaatggcttcagtggtgcagatgcacattacatttatttactactag 2858
Db 68341 TGATCCAAACAAAAATGGCTTCAGTGGTGCAGATGTCACCTACATGTTATTCTAGTACTAG 68282

Qy 2859 aaactgaagaccatgtggagacttcataacacatgggttttagtttccaccagaatggaaa 2918
Db 68281 AAACCTGAAGACCATGTGGAGACTTCATCAACATGGGTTTAGTTTTCACCAAGATGGAAA 68222

Qy 2919 gacctgtacccctttttggtggtcttactagctgggtgggtgtctgttttggagcttatt 2978
Db 68221 GACCTGTACCCCTTTTGGTGGTCTTACTAGCTGGGTGGGTGCTGCTGTTTGGAGCTTATT 68162

Qy 2979 tagagctcctagtttctcctacttataaagtagaataagtgagattgttttcttttacc 3038
Db 68161 TAGAGCTCTAGTTTCTCTACTTATAAAGTAGAAATGGTAGATTGTTTCTTTTCTTACC 68102

Qy 3039 ktaaagggagatggtaagaacaatgaatgtcttttttcaaacatttatgacaagtgtatt 3098
Db 68101 TTAAGGGAGATGGTAAGAAACAAATGAATGCTCTTTTTCAAACTTTATTGACAAAGTGATT 68042

Qy 3099 ttcaagtctgtttcaaaaaatatatttcattgtacacctgtgatccagaagaagggttccca 3158
Db 68041 TTCAAGTCTGTGTTCAAAAAATATATTTCATGTACCTGTGTGATCCAGCAAGAGGGAGTCCCA 67982

Qy 3159 gtcaagagtcactacaactgatttagttgttttagagaatgagaatggaacagttaggaat 3218
Db 67981 GTCAAGAGTCACTACAACTGATTAGTTGTTTAGAGAAATGAGAAATGGAACAGTGAAGAAAT 67922

```

Qy	3219	ggaggccaattttccatgacttcccttgtaaaacagaagcaacagaagggacaaagggctg	3278
Db	67921	GGAGGCCATATTTCCATGACTTCCCTTTGTAACAGAAAGCAACAGAAGGGACAAAGGGCTG	67862
Qy	3279	gccttcacatcactcacccttccaaactcttgtagaagtcgatctacttgcgaacccaa	3338
Db	67861	GCCTCTACATCACTCTCACCTTCCCAATCTTGTGTGAAGTGCATCTACTTGGCACAACCA	67802
Qy	3339	attaacttacttccaaagtcttggctgctgttcaggtggaaactccagctgcgaagggagttag	3398
Db	67801	ATTAACCTACTTCCAAGTCTGTGCTGTTCAGGTGGAACCTCCAGCTCAAGGGAGTTAG	67742
Qy	3399	ggaatgaaggtcttttttaaaagcttctcagccttcccttagggaacagaanaatttgggtga	3458
Db	67741	GGAAATGAAGGTCTTTTTTAAAGCTTCTCAGGCTTCCCTAGGGAACAGACATTGGGTGA	67682
Qy	3459	gccaatctgaaatttctactacaggtcattgagaccagttagatcatcttgaaatatataga	3518
Db	67681	GCCAACTGCAATTTCTACTACAGCATTTGAGACCAGTTAGATTTATTTGAAATATTATAGA	67622
Qy	3519	gagttatgaacacttaaatatgatagtggtgatgacattggatagaacatgggatacttt	3578
Db	67621	CAGTTATGAACACATTAATTTATGATAGTGGTATGACATTGGATGGATGAACATGGGATACTTT	67562
Qy	3579	agaagtgaattgacagggcaattagttgatgaaatggagtcatttgagtcctttaata	3638
Db	67561	AGAAGTAGAAATTGACAGGGCATAATTAGTTGATGAAATGGAGTCAATTTGAGTCTCTTAATA	67502
Qy	3639	gccatgatcataattaccaaagtgaagctgggtggaaacatatggtctccattttacagtta	3698
Db	67501	GCCATGTATCATANTACCAAGTGAACTGGTGGGAACATATGGTCTCAATTTACAGTTA	67442
Qy	3699	aggaataatgagacagattaatattgtytctgtcatgccacaacatccccttctbaagga	3758
Db	67441	AGGAATAATAATGGACAGATTAATATTGTCTCTGTCAATGGCCACAATCCCTTTCTAAGGA	67382
Qy	3759	agactgcctactatagcagtttttatattgttcaatttgaataataagaatgagagag	3818
Db	67381	AGACTGCCCTACTATAGCAGTTTTTATATTTGTCAATTTATGAATATTAATGAATGA -GAG	67323
Qy	3819	ttctggtacacctctgtctttacaaatatggggtgtgtccagtagtatcttccctttttaac	3878
Db	67322	TTCTGGTACCTCCCTCTTTACAAATATGG - - -TGTGTGCAGTATTTTCCITTTTAA	67267
Qy	3879	cmittcccaattccgggtgtaggtggagtgttccatttgggttttaattgtatatccct	3938
Db	67266	CCATTCCAACTCGGTGTGTAGTGATG - - -TTTCATTTTGGTTTTTAATTTGTATATCCCT	67212
Qy	3939	gatagctataattgggtcatagaatactcttatacattctagatgcgaagtccttgcgcg	3998
Db	67211	GATAGCTATAATTTGGGTCAATAGAAATTCCTTTATACATTTCTAGATGCAAGTCTCTTGTCCG	67152
Qy	3999	atatactattgagatattacacctagctctgtggcttgactgtttcttttatgctctttg	4058
Db	67151	ATATACGTATTAGATATTACACCTAGTCTGTGGCTTGACTGTTCTTCTTATGTCCTTTTG	67092
Qy	4059	atgaatagaagttttaaatttttgacaaggtcacaatttattttttcttttgttgatat	4118
Db	67091	ATGAAATAGAAGTTTTAAATTTTGAACAGGTCAAAATTTATTTTTTCTTTTGTGTGATATT	67032
Qy	4119	tttctctccaaatttaacccaagatttcagatatctctctatataataaacttata	4178
Db	67031	TTTTTCTCCAAATTTAAACCCCAAGATTTACAGATTTCTGCTCTATATTATATAACTTTATA	66972
Qy	4179	tttttatatttgtatctacaccttgaattgatatgtattgttgcaattatggatcagggtt	4238
Db	66971	TTTTTATATTTGTATCTACCTTGAAATGTATGTATGCTTGAAATATGATCAGGGTT	66912
Qy	4239	ctttttttcccccatacaagtatccagtcattgttaaactgttttattgaagaattatcc	4298
Db	66911	CTTTTTTCCCCCATACAAGTATTCACAGTCATTTGTAACATCTGTTATTTGAAAGATTATCC	66852

Qy	4299	tttctcattaaattaccttcgcgaattagtaaaaaataccaatcattm	4348
Db	66851	TTTTCTCATTTAAATACCTTGCCAATTAGTAAAAAATCAATTAACCATAA	66802
RESULT 6			
AK001017	AK001017	2044 bp	linear
LOCUS	Homo sapiens cDNA FLJ10155 fis, clone HEMBA1003433, highly similar to Homo sapiens gene for NBS1.		
DEFINITION	AK001017		
ACCESSION	AK001017.1 GI:7022031		
VERSION	Oligo capping; fis (full insert sequence).		
KEYWORDS	Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to mRNA, clone.lib:HEMBAL clone:HEMBAL1003433.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
REFERENCE	1 (bases 1 to 2044)		
AUTHORS	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sato,H., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2044)		
AUTHORS	Isogai,T. and Otsuki,T.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)		
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.		
FEATURES	Location/Qualifiers		
source	1..2044		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="HEMBAL1003433"		
	/tissue_type="whole embryo, mainly head"		
	/clone_lib="HEMBAL"		
	/dev_stage="embryo, 10 weeks"		
misc_feature	/note="cloning vector: pME18SFL3"		
	1282..1283		
	/note="118 bases segment is present in AF051334 and AF058696"		
misc_difference	1378		
	/note="compared to AF051334 and AF058696"		
BASE COUNT	723 a 384 c 442 g 495 t		
ORIGIN			
Query Match	41.7%; Score 1836.4; DB 9; Length 2044;		
Best Local Similarity	94.3%; Pred. No. 0;		
Matches 1976;	Conservative 0; Mismatches 1; Indels 119; Gaps 2;		
Qy	13	cgcgggtgacagtgcggcccaagccctgaggagcgcgacccgatgtggaactgctgccgc	72
Db	67	CGCGGTTGCAGCTCGGCCCGCCAGCCCTGTAGGAGCGCGACCGATGTGGAAACATGCTGCCGC	126
Qy	73	cgcgggcccgcagagagagaccatacacagacttttgactgcttgagtcagctgttgg	132
Db	127	CGCGGGCCCGCAGAGAGAGAACCATACAGACATTTTGACTGGCGTTGAGTACGTTGTGG	186
Qy	133	aaggaaaaactgtgcattcttaattgaaatgatcagtcgcatcgcgcgaaatcatgctgt	192

```
Db 187 AAGGAAAACTGTGCATCTCTAATGAAATGATCAGTCGATCAGCCGAAATCATGCTGT 246
Qy 193 gttactgtaactttctgttaaccaacctgagtcacaacagatgaataccctgtattgac 252
Db 247 GTTAAGTCTAATCTTCTGTACCAACCTGAGTCAACAGATGAATCCCTGTATTGAC 306
Qy 253 attaaagataattctaagtatggtacccttgttaattgaggaaaaatgcagaatggctt 312
Db 307 ATTAAGAATAAATCTAAGTATGTTGACCTTTGTTAATGAGGAAAAATGCAGAAATGGCTT 366
Qy 313 tcccgaaacttgaaagtcggggatggtattacttttggagtggttggaagtaaatccaag 372
Db 367 TTCGGAATTTGAAGTCGGGGATGGTATTACTTTTGGAGTGTGGAAAGTAAATTCAG 426
Qy 373 aatagatgatgaccttgggttgatgctctctcttctttagatgctctctgggaaactgc 432
Db 427 AATAGAGTATGAGCCTTGGTGTGATGCTCTCTCTTTAGATGCTCTGGGAAACTGC 486
Qy 433 ttaaatcaagctatatgcaacttggaggatttacttgaacaatttgacagaagaatg 492
Db 487 TTTAAATCAAGCTATTTCGAACCTTGGAGGATTTACTGTAAACAAATTCACAGAAGAATG 546
Qy 493 cactcacctgtcatggtatcagtgaaagttaccattaaacaatatgtgcaactattg 552
Db 547 CACTCACCTTGTCTATGTTGATCAGTGAAAGTACCATTAAACAATATGTGCACCTCATTTG 606
Qy 553 tggacgtccaattgtaaagccagaaatatttactgaattcctgaaagcagtttcagtcaca 612
Db 607 TGGAGCTCCAATTGTAAGCCAGANATTTTACTGAATTCCTGAAGCAGTTCAGTCCAA 666
Qy 613 gaagcagctccacaataatgaaagtgtttaccacctcttgatgaacctatttggag 672
Db 667 GAAGCAGCTCCACAAATGAAAGTTTTTACCCACTCTTGATGAACCATCTATTGGAG 726
Qy 673 taaaaatgtatctgtcagacgcaggaagaagaaacaaatcttcaaggaagaaacatt 732
Db 727 TAAAAATGTTGATCTGTCAGACGCGCAGGAAGAAACAAATCTTCAAGGGGAAACAT 786
Qy 733 tatatttttgatgcacacagcataaagaattgagttcccgagttgctcttggaggtgg 792
Db 787 TATATTTTGAATGCCAAACAGCATAAGAAATTGAGTTCGCCAGTGTGCTTTGGAGGTG 846
Qy 793 ggaagctaggttgataacagagagaatgaagaagaacataattcttcttggctccggg 852
Db 847 GGAAGCTAGGTGATAACAGAAGAGAATGAAGAACAATAATTTCTTTTGGCTCCGGG 906
Qy 853 aacgtgtgtgtgatacaggaaataacaaactcacagaccttaattctctgactgtcagaa 912
Db 907 AACGTGTGTTGATCAGGAATAACAACTCACAGACCTTAATTCCTGACTGTCAAG 966
Qy 913 gaaatggattcagtcataatggtatgctccaaaggcaagggtcttagacctatttcttga 972
Db 967 GAAATGGATTGAGTCAATAATGGATATGCTCCAAAGGCAAGGTCTTAGACCTATTCTCTGA 1026
Qy 973 agcagaatttgatggcggtgattttcatgactacaagaattactgtgactccagg 1032
Db 1027 AGCAGAAATTTGGATTGGGGGTGATTTTCATGACTACAAGAATTTACTGTGATCTCTCAGG 1086
Qy 1033 ccattccagtcagatttaaagacaacaactccaggaccaagcctttcacaggcgtgtc 1092
Db 1087 CCATCCAGTACAGATTAAAGACACAACTCCAGGACCAAGCCTTTCACAAGGCGGTGC 1146
Qy 1093 agttgatgaaaaataatgccaagcgcccccagtgaaacctacaacatcagtagctgcac 1152
Db 1147 AGTTGATGAAAAACTTAATGCGCAAGCGCCAGTGAACACTACAACATACGTAGCTGACAC 1206
Qy 1153 agaatcagacagcaatcatggttgaagtgaaagccaaagaaatacaagtcctc 1212
Db 1207 AGAATCAGACAGCAGATACATGGGATTTGAGTGAAGGCCAAAGAAATCAAAAGTCTC 1266
Qy 1213 caaatggacaaaaatttcagatgcttttcacaagacgcacccactgtaaaaggagtcctg 1272
|||||
```

```
Db 1267 CAAAATGGAAACAAAA----- 1282
Qy 1273 caaaacaagctctaataataatagtgatgatacaataactttggctaaagatgagaatccc 1332
Db 1283 ----- 1282
Qy 1333 aaactatcagctttccaccaactaaattgccaagtataataaaagtaaaagtagggcttc 1392
Db 1283 -----CACCAACTAAATTTGCCAAGTATAAATAAAAGTAAAGATAGGCCTTC 1328
Qy 1393 tcagcagcagcagacccaactccatcagaaaaactcttcaagccgtctacc-aaaaaaagg 1451
Db 1329 TCAGCAGCAGCAGACCAACTCCATCAGAACTACTTTTCAGCCGTCTACCAAAAAAGGG 1388
Qy 1452 aaagggtatgaagaaaatcaagaaaatgctctcatgcaaatcagcaagaatagaaaagtcct 1511
Db 1389 AAAGGATGAGAAATCAAGNAATGCTTCATGCAAAATCAGCAAGATAGAAAGCTCTT 1448
Qy 1512 gtctcttttagaacaacacacacctgtctacacctcatctgtggaataaaggagcagc 1571
Db 1449 GTTCTCTTTTAGACAAACACAACTGCTACACCTCATTTGTGGAAAAATAAGGAGCAGC 1508
Qy 1572 atctatctgagaatgagcctgtggacacaaaactcagacaaataacttattacagatacag 1631
Db 1509 ATCTATCTGAGAATGAGCCTGTGGACACAACTCAGACAAATAACTATTTCAGATACAG 1568
Qy 1632 atttaaatctatttgaaaaattctgcagtaaatctcatgctcagaaaaagctaaag 1691
Db 1569 ATTTAAATCTATTGTGAAAAATCTGCCAGTAAATCTCATGCTCAGAAAAAGCTTAAGAT 1628
Qy 1692 caaataaaaaagggaatggatgatgtggccatagaagatgaagtatttggaaacagttat 1751
Db 1629 CAAATAAAAAAGGGAATGGATGATGTGGCCATAGAAGATGAAGTATTGGAAACAGTTAT 1688
Qy 1752 tcaaggacacaaaacagagttagaattgattgtgaaagttcaaaacagaggagaagatg 1811
Db 1689 TCAAGGACACAAAACAGAGTTAGAAATTTGATGTGAAAGTTCAAAAACAGGAGGAAGATG 1748
Qy 1812 tcaatgttagaaaaaggccaaggatggatatagaaacaaatgacacttccagtgatgaag 1871
Db 1749 TCAATGTTAGAAAAAGGCCAAGGATGGATATAGAAACAAATGACACTTTTCAGTGATGAAG 1808
Qy 1872 cagtaccagaagtagcaaaaatatctcaagaaaaatgaaattgggaagaaacgtgaactca 1931
Db 1809 CAGTACCAGAAGTAGCAAAAATATCTCAAGAAAAATGAAATTTGGGAAGAAACGTGAACCA 1868
Qy 1932 aggaagactcactgtgtcagctaaagaaaatactcaaatgacaaacttcagatgata 1991
Db 1869 AGGAAGACTCACTATGTCAGCTAAAGAAATATCTAACAAATGGCAAACTTCAGGATGATA 1928
Qy 1992 gtgagatgcttccaaaaaagcgtgttattgactgaatttagactcactggtgattaaaaact 2051
Db 1929 GTGAGATGCTTCCAAAAAAGCTGTTATTGACTGAATTTAGATCACTGCTGATTAAAAACT 1988
Qy 2052 ctacttcagaaaatccgtctggcataaataatgatgattatggtcacaactaaaaatttc 2107
Db 1989 CTACTTCAGAAATCCGCTGTGCATAAATGATGATTATGGTCACTAAAAAATTTTC 2044
```

RESULT 7

```
AF218575 AF218575 2605 bp mRNA linear ROD 02-AUG-2000
LOCUS Rattus norvegicus Nbs1 (NBS1) mRNA, complete cds.
DEFINITION AF218575
ACCESSION AF218575
VERSION AF218575.1 GI:9651647
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2605)
AUTHORS Lanson,N.A. Jr., Egeland,D.B., Royals,B.A. and Claycomb,W.C.
```


Db 266 CAGTTTGAGTCAAAACAGATGAATTCCTACATTAAACAAATAAAAGATAAATCTCTAAGTATGG 325
Qy 277 taccttggtaagagaaataagcagaatggcttttcccgaactttgaaagtcggggga 336
Db 326 AACCTTTGTTAATGAAGAAAAAATGCAGACTGGTCTTTCTGCACGTTGAAGACAGAGA 385
Qy 337 tggattactttggagtggtttgggaagttaaaatcagaatagagatagacgtttgttgc 396
Db 386 TAGAGTTACCTTTTGGGGTGTGTTGAAAGTAATATTCAGAGTAGAATACAGAGCCCTTGGTTGT 445
Qy 397 atgctctctgtttagatgctctgagaaaaactgctttaaatcaagcttatattgcaact 456
Db 446 TTGTTCTTCTGTTAGATGTCTCGGAAAACGTGTTTAAATCAAGCTATTTTACAGCT 505
Qy 457 tggagatttactgttaacaattggacagaagaatacactcaccttgtcatggtatcagt 516
Db 506 TGGAGGACTTACCGCAAAACACTGGACGAGNAATGTACCCACCTTGTCTATGTCGCAGT 565
Qy 517 gaaagtaccattaaaaacaatatgtgcactcatttggagcgtccaatgttaaaagccaga 576
Db 566 TAAAGTCAACATTAAACCTATATGTGCACCTATTTGTCGACGTCCCAATTATATAAACCCAGA 625
Qy 577 atattttactgaactcctaagagcaattcagtcctcaagaagagcagcctccacaattgaaag 636
Db 626 ATACTTTCTGTAATTTCTCAAGCAGTTGAATCTAAGAAACAGCCCTCCAGATATTGAAAG 685
Qy 637 ttttaccacactcttgatgaaccattctattggaagttaaaatgttgatctgtcaggaag 696
Db 686 TTTTACCACCCATTGATGAACACGCTATTTGGAAGTAAAGTGTGATCTGTACGGGCG 745
Qy 697 gcaggaagaaacaaattctcaaaaggaaaaacatttatattttgaaatgccaacacga 756
Db 746 ACATGAAGAAACACAGATCTCAAGAGAAAAACATTTGTGTTTTTAAATGCCCAAGCAGCA 805
Qy 757 taagaataatgagttccogagttgcttctggaggtggggaagctaggttgataacagaaga 816
Db 806 CAAGAAACTCAGCTCGGCAAGTTGCTTTTCGGAGGTGGAGAAGCCAGGCTCATGGCAGAAGA 865
Qy 817 gaatgaagaagaacataaattcttttggctccgggaacgtgtgtgtgtgatacacagaat 876
Db 866 CGACGAAGAGGAACAGAGCTTCTTTTCAGCTCCCGGAACCTGGTTGTTGATGTAGGAAT 925
Qy 877 aacaaactcacagacotttaattcctgactgtcagaagaataatggattcagtcataatgga 936
Db 926 AACGAATACACAGCTCATTAATTCACACTCCCGAGAAAAAATGGATTCATTTGATATGGA 985
Qy 937 tatgtccaaaaggcaaggtccttagaaccttattcctgaagcagaaatggattggcgtgat 996
Db 986 TACACTTCAAGGAATGGGTCTCAGACCTATTTCTGAAGCGGAGATTGGATTTGGCTGTTAT 1045
Qy 997 ttctatgactacaagaattactgtgatcctcagggccatcccgatcacaggattaaagac 1056
Db 1046 TTTTATGACTACAGAGAATTTACTGTAATCGCAGGGCCAGCCTTGTACAGAAATTAAGAC 1105
Qy 1057 aacaaactccagaccgaagcctttccaagggcgtgctgagttgatgaaaaactaatgccaag 1116
Db 1106 AACGACTCCAGACCACCAAGCCTTTCCCAAGTCCCTGTGAGGCAATATGAAAAATATCCCAAG 1165
Qy 1117 cgccccagtgaaactacaacatcacgtagctgaacagaatacagaaacagaaacagatacatg 1176
Db 1166 TGCTCCAGTGAATATGACCCACATACCTAGCTGACACAGAATACAGAGCCAGAGATACATG 1225
Qy 1177 ---ggatttgagtgaaaggccaaaagaatacaagctctccaaaatggaaacaaaattcag 1233
Db 1226 TATGCTTTTGTAGTGAAGAACCGAAGAACTAAAGATCCCTGGACTGGAAACAAAGCTCTAG 1285
Qy 1234 aatgctttcaagacgaccccaactgttaaagagtgctctgaaacaaagctcttaataataa 1293
Db 1286 GAAACTTTCACAGAACAATTCATATATAAGGAGGGCCCTTAAACCAAGCTTCCAAAGCTAA 1345
Qy 1294 tagtatggtatacaataactttggctaagatgagaatcccaaaactatcagctttccaccaac 1353
Db 1346 CAACGTAGCATCAGATACGCTGGTTAGAGGAAAGACCCCGAGCTATCAGCTTTCTCCAAT 1405

Qy 1354 taaattgccaaagtataataaaaaagataagggtcttccagcagcagcagaccaactc 1413
Db 1406 GAAATTTCTGTTGCAAAATAAAATAAGGATTGGACTTCTCAGCAGCAGCAG ---AATC 1462
Qy 1414 catcagaactactttcagccgtctaccacaaaaaagggaaggatgaagaataatcaaga 1473
Db 1463 CATCAAAAACCTACTTCCAGCCATGCACCAAAAAGGAAAGGATGAAGACAACCCAGA 1522
Qy 1474 aatgtcttcatgcaaatcagcaagaataagaacgtcttctctcttttagaacaacaca 1533
Db 1523 GCTGTCTCTGTGCAAAATCATCCAGGATGGAGCTGTCTGTCTCTTTTGAACAACACACA 1582
Qy 1534 acctgtcacacctctattgtggaaaaataaggagcagcatctatctgagatgagcctgt 1593
Db 1583 ACCTGCGGACCCCTCACTGTGTGAAAAAGCAAG ---AGCATCAGTCTCAGAATCGCACCT 1639
Qy 1594 ggacacaaactcagacaataaactatttacagatacagatttaaaatctctttagaaaaa 1653
Db 1640 GGACAGGGAAGCGCACACCTCATCTGTGGTGGGATGGACATAGAACTCAACAGGAAGAG 1699
Qy 1654 ttctgccagtataatctcatgctcagaaaaagctaaagatacaataaaaaaagggaatgga 1713
Db 1700 TCTGTACAGGAACCACTTCCCACAGAAACTCTTTAGACC ---CAGAAAAAGAAAGATGT 1756
Qy 1714 tgaatgtggccatagaagatgaagtatttggacagttatttcaagggacacacaaaccagatt 1773
Db 1757 TGATTTAGCTACAGAAGAGAACTCTTGGAAAGATTACTCAGGAGTACAAAGCCAGAGTT 1816
Qy 1774 agaaattgatgtgaagatttcaaaaacagggaggaagtgtcaattgttagaaaaaggccaag 1833
Db 1817 GGCAGTTCAAGTTGAAGGTTGAAAAGCAGGAGCGAGATGACACCATCAGAAAAAAGCCAAAG 1876
Qy 1834 gatggatatagaacacaaatgacactttcagtgatgaagcagtagaccagaagtagcaaaat 1893
Db 1877 GATGGACGAGAGAGGAATCGGCCCTCTGAATGTGTGATCAGAGCCGGAAGAACACACGC 1936
Qy 1894 atctcaagaaaaatgaaattgggaagaaacgtgaactcaaggaagactcacatgtggtcagc 1953
Db 1937 TCCTTCAAGAAGATGAAGAGAAAGAAAGATGAAC TTCACACAGAGTCTGGTGGTCAACAAA 1996
Qy 1954 taagaataatctacaatgcacaaacttcaggatgatagtagtgatgcttccaaaaaagct 2013
Db 1997 ACATGAAATAGCTTAATAGTGTATGCTCTCAGGACAGCAGTGTAGGAGCTGCACGGAACT 2056
Qy 2014 gttattgactgaatttagatcacctggtgattaaaaactctacttccagaaatccgtctgg 2073
Db 2057 GCTGCTGACTGAATTTAGGTTCATTTGTTGTGCTAATCCACCACTCCACCTCCAGAAATCT 2116
Qy 2074 cataatgatgatgttggtcaactaaaaaatttcaagaaattcaaaaaggtccatatcc 2133
Db 2117 ATGTGTAATGAATGTGGTCCACTGAAGAAATTTCAAGAAATTTCAAGAAGCGCATTTCC 2176
Qy 2134 tggagcagaaaaacttccacacatcattggaggaatcagatctaatagctcatatgctcg 2193
Db 2177 TGGGGCAGGAAACCTTCCACACATCATCGGAGGATCAGACTTAGTAGCCACCATTGCTCG 2236
Qy 2194 aaagaatacagaacttagaagagtggttaaggcaggaatggagggtacaaaatcaacatgc 2253
Db 2237 AAAGAATACCGAGTTAGAGAATGGTTGAACACAGGAATTCGAGGTACAGAAACAACAAAGC 2296
Qy 2254 aaaagaagagctcttctgtgatgatcttttttagatacaactccttatttaaaagggaag 2313
Db 2297 AAAGGAGGAATCTCTTGTCTGATGATCTGTTTACATATATATCTTAATCTTAATAA ---AGAAG 2353
Qy 2314 ataactgaggaattttaaaaagaagccatggaaaaacttccctagtaagcatctacttcagg 2373
Db 2354 ATAAATTGAACCTTGTAAAGGAAGCTGTAGAAACACATTTTCTACTTTCAGACCAACTGAGAT 2413
Qy 2374 ccaacaggttatgaatatatagtgat 2403
Db 2414 ACATTAAATATAGTACATAGAAATATATAAAT 2443

Db 194 CAGTTTGGATCAACACAGATGAATTCCTACATTAACAAATAAAAGATAATTCCTAAGTATGG 253
Qy 277 tacotttgtaaatgagaaataatgcagaatggctttcttcgcgaactttgaaagtggggga 336
Db 254 AACCTTTGTTAATGAAGAAAAATGCAGACTGGTCTTCTTCGACGTTGAAGACAGGAGA 313
Qy 337 tggattacttttgagtggttggaaataaattcagaatagagatatagcctttgggtgc 396
Db 314 TAGAGTTACCTTTGGGGTGTGTGAAGTAATTTAGAGTAAATTCAGAGTAGAATACGAGCCCTTGGTGT 373
Qy 397 atgcctctctttagatgctctggaaaaactgctttaaatacaagctatatattgcaact 456
Db 374 TTGTCCTCTCTGTTAGATGCTCTGGAAAAACHGTGTTTAAATCAAGTATTTTACAGCT 433
Qy 457 tggaggatttactgtaacaattggacagaagaatgcactcaccttgcattgcatggtatcagt 516
Db 434 TGGAGGACTTACCCCAAACTGGACGGAAGAATGTACCCACCTTGTCTCATGTGCGCAGT 493
Qy 517 gaaagttacaaataatataatgacactcatttggacactatttggacgtccaattgtaagccaga 576
Db 494 TAAAGTCAACCAATTAATACTATATGTGACTCATTTGTGGACGTCCAATTTATAAAACCCAGA 553
Qy 577 atatttactgaattcctgaaagcagtttcagtcacaagaagcagcctccacaaattgaaag 636
Db 554 ATACTTTTCTGAATTTCTCAAGCAGTTGAATCTTAAGAAACAGCCTCCAGATATTGAAG 613
Qy 637 ttttaccacaccttgcataacattctattggaagtaaaatgttgatctgtcaggagc 696
Db 614 TTTTACCACCCATTGATGAACCACTATTGGGAAGTAAAGTCTTGATCTGTCAAGGCG 673
Qy 697 gcaggaaagaaacaaatcttcaaaaggaaacaaatktatatatttgaatgccaaacagca 756
Db 674 ACATGAAGAAAAACAGATCTTCAAGGAAAAACATTTGTGTGTTTAAATGCCAAAGCAGCA 733
Qy 757 taqaaattagttccagttgcttcttgaggtgggaagcagtggtgataacagaga 816
Db 734 CAAGAACTCAGCTCGGCAGTTGCTTTCGGAGGTGGAGAGCCAGGCTGATGGCAGAGA 793
Qy 817 gaa tgaagaagaacataattcttttggctccgggaacgtgtgtgttgatagacaggaat 876
Db 794 CGAGGAAGGGAACAGAGCTCTTTTCAGCTCCCGGAACCTGCGCTTGTGATGATGAAGAT 853
Qy 877 acaaaactcacagaccttaattcctgactgcagaagaataatggattcagtcocataatggga 936
Db 854 AACGAATACACAGCTCATATAATTTCCACACTCCCAAGAAAAATGGATTCATTTGATAATGGA 913
Qy 937 tatgctccaaggaaggtctagacctattcctgagcagaaattgagttggcgtgat 996
Db 914 TACACTTCAAGGAATGGTCTCAGACCTATTCTTGAAGCGGAGATTGGATTGGCTGTAT 973
Qy 997 ttctagctacaaagaattactgtgatcctcagggccatccccagtcacaggaattaaagac 1056
Db 974 TTTTATGACTACAGAGAATTACTGTAATCCGACGGCCAGCCTTTGTACAGAAATTAAGAC 1033
Qy 1057 acaactccaggaccagcctttcacaggcgtgtcagttgatgaaaaactaatggccaaag 1116
Db 1034 AACGACTCCAGGACCAAGCCCTTTCCCAAGTCTCTGTACGCCAATGGAAAAATAATCCCAAG 1093
Qy 1117 cgccccagtgaaacactacaatactagctgacacagaaatcgagcaagcaagacatacatg 1176
Db 1094 TGCTCCAGTGAATATGACCACATACGTAGCTGAGACAGAAATCAGAGCCAGCAGATACATG 1153
Qy 1177 ---ggatttgtagtgaaggccaaagaaatacaagttcccaaatggaaacaaatttcag 1233
Db 1154 TATGCCTTTGTAGTGAAGACCAAGGAAGTAAAGATCCCTGGACTGGGAACAAAGCTCTAG 1213
Qy 1234 aatgctttcaagagcagccacctgtaaaggagtcctgcgaacaaagcctctataataa 1293
Db 1214 GAAACTTTTCAAGAAACATTTCAATATATAAGGAGGCCCTTAAACCAAGCTCCAAAGCTAA 1273
Qy 1294 tagtatggtatcaaaacttctggctgaagatgagaatcccaactcatcagctttcaccaac 1353
Db 1274 CAAGGTAGCATCAGTACGTGGTTAGAGGAAGACCCCGAGCTATCAGCTTCTCCAAAT 1333

Qy 1354 taaattgccagtgataataaaaaagtaaaagtagggcttctcagcagcagcagaccaactc 1413
Db 1334 GAAATTTCTTGTCAAATAAAAAATAGGATTGGACTTCTCAGCAGCAGCAG---AACTC 1390
Qy 1414 catcagaaactacttttcagccgtctcacaacaaaggaaggaaggaataaacaaga 1473
Db 1391 CTTCAAAACTTACTTTCAGCCATGCACGAAAAAGGGAAGGATGAAGACACCCAGA 1450
Qy 1474 aatgtcttcagcaaatcagcaagaatagaaacgtctgttctctcttttagaacaacaca 1533
Db 1451 GCTGCTCTGTCGAATAATCATCCAGATGGAGCTGCTCTGTCTCTTTTAGAACAAACACA 1510
Qy 1534 acctgtcacacctcattgttgaaaaataaaggacagcatctatcttgagaatgagcctgt 1593
Db 1511 ACCTGCCGAGCCCTCACTGTGGAAGCAAG---AGCATCAGTCTCAGAAATGGACCCCT 1567
Qy 1594 ggacacaaactcagacaataaacttatttacagatcacagatttaaaatctatttgaaaaa 1653
Db 1568 GGACAGGGAAGCCGACACCTCATCTGGTGGGATGGACATGAACCTCAACAGAGAGAG 1627
Qy 1654 tctgccagtaaatctcatgctgcagaaaagctaagatcaaatataaaaaagggaatgga 1713
Db 1628 TCTTGACAGGAACCACTTCCCACAGAAACTCTTAGACC---CAGAAAAAAGAAAGATGT 1684
Qy 1714 tgatgtggccataagaagatgaagtattggaacagttatttcaaggagcacacaaaccaggtt 1773
Db 1685 TGATTTAGCTACACAGAGGAAGTCTTGGAAAGAGTTACTGAGGAGTACAAAGCCAGAGTT 1744
Qy 1774 agaattgagtgaagtctcaaaacagggagggaagtgcataatcttagaaaaagcccaag 1833
Db 1745 GCGAGTTTCAAGTGAAGGTTGAAAAAGCAGGAGCAGATGACACCATCAGAAAAAGCCCAAG 1804
Qy 1834 gatgatatagaacaaatgacactttcagtgatgaagcagtcacacaaatagcaaaat 1893
Db 1805 GATGGACGAGAGAGGAATCGGCCCTCGAATGGTGGATCAGAGCGCGGAAAGCAACAGCGC 1864
Qy 1894 atctcaagaaaaatgaattgggaagaacgtgaactcaaggaagcactcactatggtcagc 1953
Db 1865 TCITCAAGAGATGAAGAGAAAAAGAAAGATGAACCTTCAGACACAGATCGTGGTCAACAAA 1924
Qy 1954 taagaataatctaaacaaatgcacaaactccaggaatgtagtgagatgcttccaaaaagct 2013
Db 1925 ACATGAAATAGCTAATAGTGTGCTCTCAGGACAGCAGTGAGGAGGCTGCCACGGAACCT 1984
Qy 2014 gtattgactgaatttagatcactggtgattaaaactctacttccagaaaaatccgtctcg 2073
Db 1985 GCTGCTCAGTGAATTTAGGTCATTTGGTTGTGTCAGTAACCAACACTCCACCTCCAGAAATCT 2044
Qy 2074 cataaattgatgattatggtccaactaaaaatttcaagaaattcaaaaaggtcacatatcc 2133
Db 2045 ATGTGTAATGAATGTGCTCCACTGGAAGATTTCAAGAAATTTCAAGAGGGGACATTTCC 2104
Qy 2134 tgaagcagggaaaaacttccacacatcattggaggatcagatcataatagctcatgctcg 2193
Db 2105 TGGGGCAGGAAAGCTTCCACACATCATCGGAGGATGAGACTTAGTAGGCCACCATGTCTG 2164
Qy 2194 aagaatcacagaactagaagatggctcaaggcaggaatggaggtacaaatcaaatcagc 2253
Db 2165 AAAGAATCCGAGTTAGAAGAAATGTTGAACACAGAAATGGAGGTACAGAAACAAACAGC 2224
Qy 2254 aaaaagaagtcctctgctgatgattcttttagatacaatccttattttaaaagagaag 2313
Db 2225 AAAGGAGGAATCTCTTGTCTGATGATCTGTTTAGATATAATCTTAATGTAAAA---AGAAG 2281
Qy 2314 ataactgaggtatttaaaaaagaagccatggaaaaacttcttagtaagcatctacttcagg 2373
Db 2282 ATAACTGAACCTTGTAAAGGAAGCTGTAGAAACACATTTCTTACTTCAGACCACTGAGAT 2341
Qy 2374 ccaacaaggttatataatgaatatagtgat 2403
Db 2342 ACATTATATAGTACATAGAAATTTATAAT 2371

[illegible]

* 30247	gap of unknown length				
* 30973:	contig of 727 bp in length				
* 30974	gap of unknown length				
* 31721:	contig of 748 bp in length				
* 31722	gap of unknown length				
* 32456:	contig of 735 bp in length				
* 32457	gap of unknown length				
* 33185:	contig of 729 bp in length				
* 33186	gap of unknown length				
* 33923:	contig of 738 bp in length				
* 33924	gap of unknown length				
* 34648:	contig of 725 bp in length				
* 34649	gap of unknown length				
* 35376:	contig of 728 bp in length				
* 35377	gap of unknown length				
* 36107:	contig of 731 bp in length				
* 36108	gap of unknown length				
* 36832:	contig of 725 bp in length				
* 36833	gap of unknown length				
* 37589:	contig of 757 bp in length				
* 37590	gap of unknown length				
* 38342:	contig of 753 bp in length				
* 38343	gap of unknown length				
* 39092:	contig of 750 bp in length				
* 39093	gap of unknown length				
* 39838:	contig of 746 bp in length				
* 39839	gap of unknown length				
* 40586:	contig of 748 bp in length				
* 40587	gap of unknown length				
* 41330:	contig of 744 bp in length				
* 41331	gap of unknown length				
* 42108:	contig of 778 bp in length				
* 42109	gap of unknown length				
* 42824:	contig of 716 bp in length				
* 42825	gap of unknown length				
* 43539:	contig of 715 bp in length				
* 43540	gap of unknown length				
* 44280:	contig of 741 bp in length				
* 44281	gap of unknown length				
* 45022:	contig of 742 bp in length				
* 45023	gap of unknown length				
* 45775:	contig of 753 bp in length				
* 45776	gap of unknown length				
* 46518:	contig of 743 bp in length				
* 46519	gap of unknown length				
* 47266:	contig of 748 bp in length				
* 47267	gap of unknown length				
* 48011:	contig of 745 bp in length				
* 48012	gap of unknown length				
* 48758:	contig of 747 bp in length				
* 48759	gap of unknown length				
* 49505:	contig of 747 bp in length				
* 49506	gap of unknown length				
* 50253:	contig of 748 bp in length				
* 50254	gap of unknown length				
* 50998:	contig of 745 bp in length				
* 50999	gap of unknown length				
Query Match	15.7%;	Score 692.8;	DB 2;	Length 64310;	
Best Local Similarity	96.2%;	Pred. No. 1.2e-114;			
Matches 717:	Conservative	1;	Mismatches	26;	Indels 1; Gaps 1;
Qy 2774	tgctgtagcttaactaaactctggtgatccaacaaatggcttcagtggtgcagatgt	2833			
Db 60577	TGCTGAATCCGCTCTTACTACTGTANAGNANACACAATGGCTTCAGTGGTGCAGATGT	60636			
Qy 2834	cacctacatgttattctagtagtaagaactgaagaccatggagacttcaataaagcatg	2893			
Db 60637	CACCTACATGTTATTCCTAGTACTAGAACTGAAGA-CATGTGGAGACTTCATCAACATG	60695			
Qy 2894	ggtttagtttcaccagaaatggaagaccctgtacccttttctggtcttactgagctg	2953			
Db 60696	GTTTAGTTTCCACAGAAAGGAGACCTGTACCCCTTTTGGTGGTCTTACTGAGCTG	60755			

Qy 2954	ggtgggtgtctgttttgagcttatttagagtccttagtttctactactataaagtgaagt	3013			
Db 60756	GGTGGGTGTCTGTTTGGAGCTTATTTAGAGTCCTAGTGTCTTCTACTTATAAGTAGAAT	60815			
Qy 3014	ggtgagattgttttcttttttttttacccktaagaggagatagtaagaacaataaagtctttt	3073			
Db 60816	GGTGAGATTGTTTCTTTTCTTACCTTAAGGGAGATGGTAAGAAACAATGAATGTCCTT	60875			
Qy 3074	tttcaaacctttatgacaagtatttcaagtcgtgttcaaaaatatatcatgtacct	3133			
Db 60876	TTTCAAACTTTATTGACAAAGTATTTCAGTCTGTCTTCAAAAATATATTTCATGTACT	60935			
Qy 3134	gtgatccagcaagaaggagttccagtcgaagtcactacactgattagttggttagag	3193			
Db 60936	GTGATCCAGCAAGAAGGAGTTCAGTCAAGAGTCACATGATTAGTTGTTTACAG	60995			
Qy 3194	aatgagaaatggaacagtggaggaatggaggccatatattccatgactccctctgtaaacag	3253			
Db 60996	AATGAGAAATGGAACAGTGAAGGAATGGAGGCATATTTCCATGACTTCCCTTGTAAACAG	61055			
Qy 3254	aagcaacagaaaggacagagctggcctctacatcactctcaccctccaaatctgtgg	3313			
Db 61056	AAGCAACAGAAAGGACAGAGGCTGGCCTCTACATCCTCTCACCTTCCAAATCTGTGG	61115			
Qy 3314	aagtgcattctactgccagaacaaattaaacttacttccaaagtctctggtcgtgctgaggt	3373			
Db 61116	AAGTGCATCTACTTCCAGAACCAAAATTAACCTTACTTCCAAAGTCTCTGCTGCAGGT	61175			
Qy 3374	ggaactccagctgcaaggagtagggaaatgaaggctcttttttaaaagcttccagcc	3433			
Db 61176	GGAATCCAGCTGCAAGGGAGTTAGGGAAATGAAGGCTTTTTTTAAAGGTTCTTCAGCC	61235			
Qy 3434	ttcttagggaacagaaattggtgagccaatctgcaattctctactacagcattgagacc	3493			
Db 61236	TTCTTAGGGAACAGACATTTGGTGAGCCCATCTGGCATTTCTACTACAGGCTTGAGACC	61295			
Qy 3494	agttagattattgaaatattataga	3518			
Db 61296	AGTTAGATTATTGAAATATTATAGA	61320			
RESULT 12					
AF230342					
LOCUS	AF230342	2533 bp	mrna	linear	VRT 09-JAN-2001
DEFINITION	Gallus gallus Nijmegen breakage syndrome 1 (NBS1) mRNA, complete cds.				
ACCESSION	AF230342				
VERSION	AF230342.1	GI:12056575			
KEYWORDS					
SOURCE	chicken.				
ORGANISM	Gallus gallus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Gallus; Phasianinae; Gallus.				
AUTHORS	Tauchi,H., Kobayashi,J., Morishima,Ki., Matsuura,S., Nakamura,A., Shiraiishi,T., Ito,E., Masnada,D., Delia,D. and Komatsu,K.				
TITLE	The Forkhead-associated Domain of NBS1 Is Essential for Nuclear Foci Formation after Irradiation but Not Essential for hRAD50.hMRE11.NBS1 Complex DNA Repair Activity				
JOURNAL	J. Biol. Chem. 276 (1), 12-15 (2001)				
PUBLISHED	11062235				
REFERENCE	2 (bases 1 to 2533)				
AUTHORS	Tauchi,H., Matsuura,S., Takeda,S. and Komatsu,K.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-FEB-2000) Department of Radiation Biology, Research Institute for Radiation Biology and Medicine, Hiroshima University, Kasumi 1-2-3, Minami-ku, Hiroshima 734-8553, Japan				
FEATURES	Location/Qualifiers				
source	1. .2533				
	/organism="Gallus gallus"				
	/db_xref="taxon:9031"				

Db 1492 TACAAATTACTTCCATGTAGCTAGG---AAAAGGAAAGAGCTGAAGAAAGGAGAGAAAC 1548
Qy 1477 gtcttcgcaaatcagcaagaatag---aaacgtctgttctctttagaacaacacaca 1533
Db 1549 ATCTCTATCCAAACAGCAAACTGGAGAAAAGCCATTGCCTGTTTCCGAGTCCACGA 1608
Qy 1534 acctgtcacacctctatgttggaataaaggagcagcatctatctcgtgagaaatgagcctgt 1593
Db 1609 ATCTCAGCTTCATCAGCATGGATAGTGAAGAAAGAACAGC---ATGGAAGGGAATA 1664
Qy 1594 ggacacaaactcagacaataacttatattacagatacagatttaaaatctattgtgaaaaa 1653
Db 1665 ACATCCAGCTTGGAGGAAAGTGGTGAACCTGGCAAGTGATAAAACATGACATTAATA 1724
Qy 1654 ttctgcccagtaaatctctatcgtcgcagaaaagcctaagatcaataaaaaagggaatgga 1713
Db 1725 CTTTTAGTGAATC-----CTGCACCAAGAAAGAAAGAGTGTAGATGTATCTGA 1779
Qy 1714 tgatgtgcccataagaatgaagtattggaacagdtattcaaggacacacaaaaccagagtt 1773
Db 1780 AGATGTAGACGCTAGAGATGGTGTGAAAGCAGAGACCTAGACTGGGAAGGACAAAC 1839
Qy 1774 agaaatgatgtgaa-----agttcaaaaacagaggagaagatgtcaatgtttagaaaa 1825
Db 1840 AGCAAAATGGTGACAGGAAGCTCAAACCAATAAAGCAAGAAAGAGATGCTCTGGAACCAA 1899
Qy 1826 aggccaaaggtggratagaaaacaaatgacactttcagttgatgaagcagcaccagaaagt 1885
Db 1900 AGGCAGCAGAAACGGAAGGAATAACAAAGCAGAGAGAGAGAAATGAATGCTGAGAAA 1959
Qy 1886 agcaaaatatctcaagaaaaatgaaattggga-agaaaagctgaaactcaagaaagactcaact 1944
Db 1960 AGAAGAAAGTTGGATCAGTTCTGACTCTGGAAGACAAAAGAAAGATCAAGGAAGATCATC 2019
Qy 1945 atggtc---agctaaagaataatacaaatgacaaacttcagttgatgaagcagcaccagagatgct 2001
Db 2020 AGTCTCGATAAGAAACAAATTAATTAACCAACAAACAAATTTGAAGATGACTCCAGCCGTCT 2079
Qy 2002 tccaaaaagctgttattgactgaatttagatcactcgtgtgattaaaaactctacttccag 2061
Db 2080 TCCTAGCAAACTTCTGTTGACTGAGTTTCGATCACTGGTTCTCAGCTGTCCAAGATCAAA 2139
Qy 2062 aaatccgtctcgtgcaataatgatgattatgttcaactaaaaaatttcaagaaattcaaaaa 2121
Db 2140 CAGCCCAACTATGAGAATAACCAATGCAGGGGAGAGAAACAAATTTCAAAAGCTTCAGAAA 2199
Qy 2122 ggtcacatatctcgtgagcagaaaactccacacatcattggagatcagatcctaagc 2181
Db 2200 GGTCCCTTACCCAGGGCAGACAACTTCCATACATAAATTTGGAGGATCAGATTTGGTTC 2259
Qy 2182 tcatctgctcgaagaatacagaaactagaaagtggtgtaaggcagggaatggaggatca 2241
Db 2260 TCACCAAGCTAGAAAGAAATTCAGAACTAGAAAGATGGTTAAGAGAAGAAATGGAGAGCA 2319
Qy 2242 aaatcaacatgcgaagaagagtcctcgtgtgatgtatcttttttagatcacatccttattt 2301
Db 2320 GAATTCGACGTGCAAGAGAAGAAATCGCTGTGTGATGATCTCTTTAGATATGATCCTAATGT 2379
Qy 2302 aaaagagagaagataaactgaggtatttaaaaagagccat 2341
Db 2380 GAAGAGGAGAAGATAAATTTGTGACCTGAAATTTACAGCAAT 2419

RESULT 13
AX192776/c
LOCUS AX192776 646 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 343 from Patent WO0149716.
ACCESSION AX192776
VERSION AX192776.1 GI:15210732
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 646)
AUTHORS Xu,J., Lodes,M.J., Secret,H., Benson,D.R., Meagher,M.J.,
Stolk,J.A., King,G.E., Wang,T. and Jiang,Y.
TITLE Compounds for immunotherapy and diagnosis of colon cancer and
methods for their use
JOURNAL Patent: WO 0149716-A 343 12-JUL-2001;
CORIXA CORPORATION (US)
FEATURES Location/Qualifiers
source 1..646
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 239 a 124 c 95 g 188 t
ORIGIN

Query Match 12.5%; Score 552.2; DB 6; Length 646;
Best Local Similarity 93.0%; Pred. No. 2.9e-89;
Matches 609; Conservative 4; Mismatches 32; Indels 10; Gaps 3;

Qy 3419 aaaagctctcagccttccttagggaacagaaaattgggtgagcccaatctgcaatttctact 3478
Db 645 AAAGCTTTCAGCCTTCCTAGGGAACAGAAATTTGGGAGGCCCAATCTGCAATTTCTACT 586
Qy 3479 acagggcattgagaccagtttagatttgaataattatagagagttatgaacacttaatt 3538
Db 585 ACAGGCATTTGAGACCAGTTAGATTATTGAAATATTATAGAGAGTTATGAACACTTAAATT 526
Qy 3539 atgatagtggtatgacattggatgagaacatgggatacttttagaagtagaattgcagggc 3598
Db 525 ATGATAGTGTATGACATTTGGATAGAACATGGGATACCTTTAGAAAGTGAATTTGACGGC 466
Qy 3599 atattagttgatgaaatggagtcatttgagtcctytaaagccatgatacctaatacca 3658
Db 465 ATATTAGTTGTAATGAATGGAGTCAATTTGAGTCTCTTAATAGCCATGATCATATACCA 406
Qy 3659 agtgaagctggtggaacatatggtctccattttacagttaagggaataataatggagcagatt 3718
Db 405 AGTGAAGCTGGTGGAAACATATGGTCTCCATTTTACAGTTAAGGAATATAATGGACAGATT 346
Qy 3719 aatatatttctgtctcagcccaaatcccttcttaaggaagactgccctactatagcag 3778
Db 345 AATATTGTTCTGTCTGATGCCCCCAATCCCTTTCTAAGGAAGACTGCCCTACTATAGCAG 286
Qy 3779 ttttattattgtcaattttatgaataataatgaatgagagttctcgttacctcctgtcttt 3838
Db 285 TTTTATATTGTCAATTTATGAATAATAATGAATGA-GAGTTCTGTTACCTCCTGCTCTT 227
Qy 3839 acaaatattgggtgtgtgtccagtatatttcccttttaacommtecccaattccgggtgtgt 3898
Db 226 ACAAAATATGG----TGTGTGTCAGTATTTTTCCTTTTAAACCAATCCAATCGGTGTGTAG 171
Qy 3899 aggtggatgtttccatttgggttttaatttgtatatccctgataagctatattgggtcat 3958
Db 170 TGAG----TTTCATTTTGGTTTTTAATTTGTATATCCCTGATAGCTATATAATGGGTAT 116
Qy 3959 agaaattctttatcacattctagatgcagtcctcttgcggatatacgtattgagatatta 4018
Db 115 AGAAATCTTTATACATTCATAGATGCAATCTCTCTGTCGGATATATGTATTGAGATATTA 56
Qy 4019 cacctagtgtgggtgactgttttcttttatgtcttttggatgaatagaagtttt 4073
Db 55 CACCTAGTCTGTGGCTTGACTGTTTTTCTTTTATGTCTTTTGTGATGAATAGAAGTTTT 1

RESULT 14
AX341490
LOCUS AX341490 544 bp DNA linear PAT 10-JAN-2002
DEFINITION Sequence 1737 from Patent WO0196388.
ACCESSION AX341490
VERSION AX341490.1 GI:18137472
KEYWORDS human.


```
/evidence=experimental
3289. .3423
/rpt_family="L1MC/D"
complement(join(4839. .5019,6434. .6483,12661. .12774,
15548. .15703,17232. .17300,22652. .23099,24691. .24963,
28133. .28262,33818. .33915,39772. .39965,40581. .40998,
47628. .47731,50142. .50301,50783. .50931,52130. .52263,
53933. .54020))
/genes="NBS1"
/notes="Nibrin, 2.4kb transcript"
/evidence=experimental
complement(4854. .4859)
/genes="NBS1"
/notes="NBS1, Nibrin, 2.4kb transcript"
/evidence=experimental
complement(join(4989. .5019,6434. .6483,12661. .12774,
15548. .15703,17232. .17300,22652. .23099,24691. .24963,
28133. .28262,33818. .33915,39772. .39965,40581. .40998,
47628. .47731,50142. .50301,50783. .50931,52130. .52263,
53933. .53969))
/genes="NBS1"
/codon_start=1
/evidence=not_experimental
/product="Nibrin"
/protein_id="AAD08722.1"
/db_xref="GI:3108229"
/translation="MKMLPAAGPAGGEPYRLLTGVYVYVGRKNCAILLIENDQSIERN
HAYLTANFVNLSDDEIPVLTKDNSKYGFVNEEKWONGFSRTLKSGDGTGFGVF
GSKFRTEYPLVACSLDVSQKLTALNQATLOIGGTVNNWTECHLMVSVKVIK
TICALICGRPIVKPEYFTEFLKAVESKKQPPQIESYPLDPSIGSKNDLSGRBR
KQIFKGTIFILNAKHKLSSAVFVGGEARLITENEENHFFLAGPFCVVDGTGIT
NSQTLIDQCKWQISIMDLQROGLRPIPEADIEGLAIVPMFTTKYCDPQGHSTGLK
TTTTGSLGCVSVDEKLPSPAPVNTTYYVADPEQADTDLSEPKETKVKSMQK
FRLMSQDAPTVKESCTSSNNNSWNTLAKMRIPNQLSPTKLPSPINKSKDRASQK
QTNIRNYFQPTSKRDERDENQMSCKSARLETSCSLEEQPATPSLWKNKEQHL
SENEPVDNSDNLFTDLDLSIVKNSAKSHAAKLRSKKREMDQVALEDYVQL
FKTKTELEDVYKQBEDVNVKRRPRMDIETNDFSDAVPESSKISOENIEGKRR
ELKEDSLWSAKELISNDKLODDSEMLPKLLLTFLRSLVINKTSRNPSCINDDYGL
KNPKFKFTYVPGAGLPHIIGGSDLIHARKNTELEWLMQEMEYQNHAKESLAA
DDLFRYNPLYKRRR"
complement(5495. .5758)
/rpt_family="Alusg"
7018. .7204
/rpt_family="L2"
complement(7385. .7435)
/rpt_family="L2"
7620. .7799
/rpt_family="MER6B"
complement(7891. .8070)
/rpt_family="MIR"
complement(9507. .9828)
/rpt_family="AluJb"
complement(10262. .10401)
/rpt_family="MIR"
12006. .12320
/rpt_family="L1MEC"
complement(13497. .13554)
/rpt_family="AluJb"
13555. .14039
/rpt_family="MER84"
complement(14040. .14287)
/rpt_family="AluJb"
complement(16605. .16705)
/rpt_family="MIR"
16756. .17038
/rpt_family="AluSp"
17039. .17065
/rpt_family="(TAA)n"
17357. .17391
/rpt_family="U2"
18286. .18461
/rpt_family="L2"
18577. .18720
```

```
repeat_region
/rpt_family="MIR"
19368. .19467
/rpt_family="L1MD1"
19606. .19938
/rpt_family="MER2"
20473. .20739
/rpt_family="AluY"
21689. .21772
/rpt_family="MER5A"
complement(21710. .21800)
/rpt_family="MER5A"
complement(24179. .24388)
/rpt_family="L2"
26114. .26192
/rpt_family="MER58A"
26204. .27207
/rpt_family="Tigger3b"
complement(27208. .27713)
/rpt_family="MERIA"
27617. .104987
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="284N21"
27714. .27838
/rpt_family="Tigger3b"
27863. .27988
/rpt_family="MER58A"
28879. .29181
/rpt_family="AluJb"
29184. .29341
/rpt_family="MIR"
29364. .29487
/rpt_family="L2"
29780. .30050
/rpt_family="AluJo"
30058. .30077
/rpt_family="(TAA)n"
30123. .30245
/rpt_family="L2"
32351. .32517
/rpt_family="MIR"
33083. .33362
/rpt_family="L2"
33351. .33683
/rpt_family="L2"
```

Query Match 10.2%; Score 451.2; DB 9; Length 330250;
Best Local Similarity 94.1%; Pred. No. 2.2e-71;
Matches 2024; Conservative 9; Mismatches 26; Indels 91; Gaps 6;

```
QY 2278 tcttttagatacaatccttatttataaaaggagaagataactgaggattttataaagaag 2337
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 5028 TCITTCAGATACACTCTTATTATAAAAGGAGAGATACTGAGGATTTAAAGAAG 4969
QY 2338 ccatggaaaaacttcctagtagcatctacttcaggccacaagaaggttatatgaatatata 2397
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4968 CCATGGAAAAAATTCCTTAGTAGCATCTACTTCAGGGCCACAAGGTATATGAATATATA 4909
QY 2398 gtgtatagaagcatttaagttacaatgtttttatggcctaataattattataaataatgca 2457
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4908 GTGTATAGAAGCGATTTAAGTTACAATGTTTTATGGGCTAAATTTATTAATAAATGCA 4849
QY 2458 caaaactttgattctttttagtatgtaacaattgtttgtcgttttcagggtttgtcattg 2517
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4848 CAAACTTTGATCTTTTGTATGTATGAACAATTTGTTCTGTTTCAGGCTTTGTCATTG 4789
QY 2518 catcttttttcatatttttaaatgtgtttttttattataaatagtttaatatagtcagttc 2577
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4788 CATCTTTTTCATTTTAAATGTGTTTGTATTAAATAGTTAAATATAGTACAGTTC 4729
QY 2578 aaaaattcctaattacgttaaggtaagactaaagctaacctccaccattccaccattgcttagct 2637
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4728 AAAATTCTAAATGTACGTAAAGGTAAA-GACTAAAAGTCACCCCTCCACCATTTGCTAGCT 4670
```

Qy	2638	act-----	2640
Db	4669	ACTTGGTTCOCCTCAGAAAAAATTCATGACTCAATTCCTTATGAATCTTTCACGGGAT	4610
Qy	2641	-----tattttaataattctacacaataatgatacgaataac	2678
Db	4609	TTTTTGAGTCCTATTCAAATTCCTTTTTTAAATAATTTCCCTACACAATGATAGCATAAC	4550
Qy	2679	atagcagtggttcacacctctgttttacttagtaagaattaaaaattatagaataatc	2738
Db	4549	ATATGCAGTGTCTACACCTCTGCTTTTTTACTTAGT-AGATTAAAAATTTATAGGAATATC	4491
Qy	2739	aaataatgtttttaataatatttttttttttccaatatgctgtagtcttaoctaaactctgg	2798
Db	4490	AAATATAATGTTTTTAATAATTTTTTCTTTTCCATTATGCTGTAGTCTTTACCTAAACTCTGG	4431
Qy	2799	tgatccaaacaataatggcttcagtggtgcagatgtcgaactacatggttatcttagtactag	2858
Db	4430	TGATCCAAACAAATGGCTTCAGTGGTCAGATGTCACCTACATGTTATTCTTAGTACTAG	4371
Qy	2859	aaactgaagacaatgtggagacattcatcaaacatgggttttagttttccaccagaatggaaa	2918
Db	4370	AAACTGAAGACCATTGSGAGACTTCATCAACATGGGTTTAGTTTTCCACAGAAATGGAAA	4311
Qy	2919	gaactgtacccttttgtgtcttactgaactgagctgggtgggtgctctgttttgacittatt	2978
Db	4310	GACCTGTACCCTTTTGGTGGTCTTTACTGAGCTGGGTGGGTGCTGTTTGTAGCTTATT	4251
Qy	2979	tagagtcttagtttttccactactataaagttagaaaatggtagatattgtttcttttcttacc	3038
Db	4250	TAGAGTCTTAGTTTCTTACTTATAAGGTAGAAATGGTGAGATGTTTCTTTTCTTACC	4191
Qy	3039	ktaaagggagatggtaagaacaatgaatgtcttttttccaaactttattgacaagtgatt	3098
Db	4190	TTAAAGGGAGATGGTAAAGAAACAATGAATGTCTTTTTTCAACCTTTATTGACAAGTGATT	4131
Qy	3099	tcaagctgtgttcaaaaatatattcatgtacctgtgatccagcaagaaggagattcca	3158
Db	4130	TTCAAGTCTGTGTTCAAAAATATATTATGTATCTACCTGTGATCCACAAAGAGGAGTTCCA	4071
Qy	3159	gtcaagagtccaactacaactgattgtgttttagaagaatgagaaaatggaaacagtgaggaaat	3218
Db	4070	GTCAAAGAGTCACCTACAACTGATTAGTTGTTTAGAGAAATGAGAAATGCAACAGTGGGAAT	4011
Qy	3219	ggaggccatatattccatgacttcccttgtaaacagaagcaacaagaagggacaagagctg	3278
Db	4010	GGAGGGCATATTTCCATGACTTTCCCTTGTAAACAGAAGCAACAGAAGGGGCAAGAGGCTG	3951
Qy	3279	gccttcacatcactctcaccttccaaactctgtggaagtgcattctacttgcacagaaccaa	3338
Db	3950	GCCTCFACATCACTCTCACCTTCCAAATCTTGTGGAACTGGCATCTACTTGCCAGAACCA	3891
Qy	3339	attaacttacttccaagttctggctgcttgpaggtggaaotcccagctgcgaaggagttag	3398
Db	3890	ATTAACCTTACTTCCAAAGTTCCTGGCTGCTTGCAGGTGGAACTCCAGCTGCAAGGAGTAG	3831
Qy	3399	gaaaatgaagctctttttttaaagcttctcagccttctaagggaacagaaaattgggtga	3458
Db	3830	GGAAATGAAGGTCTTTTTTAAAGCTTCTCAGCCTTCCCTAGGGAACAGAAATGGGTGA	3771
Qy	3459	gccaaatcgaaatttctactacagcattgacacagttagatttatigaaatattataga	3518
Db	3770	GCCAACTGCANNTCTCTACTACAGGCATTGAGCCAGTTAGATTATTGAAATATTATAGA	3711
Qy	3519	gagttatgaacacttaaatattatgtagtggatgacattggataggacaatgggatacttt	3578
Db	3710	GAGTTATGAACACTTAAATTTATGATAGTGGTATGACATTTGGATAGACATGGGACTTT	3651
Qy	3579	agaagtgaattgacagggcatattagttgtagaaatggagctatttgacttcttctaata	3638
Db	3650	AGAAGTAGAATTACAGAGGCATATTAGTTGATGAATGGAGTCAATTTGAGTCTCTTAATA	3591

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 01:40:40 ; Search time 66.52 Seconds
(without alignments)
1960.887 Million cell updates/sec

Title: US-09-837-602-2

Perfect score: 3899

Sequence: 1 MWLLPAAGPAGGEPYRLTF.....KEESLADDLFRYPYLKRRR 754

Scoring table:

BLOSUM62

Gapop: 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3896	99.9	754	4 O60934	O60934 homo sapien
2	2710	69.5	751	11 Q9R207	Q9R207 mus musculus
3	2702	69.4	751	11 Q88981	Q88981 mus musculus
4	2702	69.3	751	11 Q9R1X1	Q9R1X1 mus musculus
5	2685.5	68.9	750	11 Q9JIL9	Q9JIL9 rattus norv
6	1766.5	45.3	753	13 Q9DE07	Q9DE07 gallus gall
7	473	12.1	811	5 Q9VT40	Q9VT40 drosophila
8	227.5	5.8	1460	4 Q75058	Q75058 homo sapien
9	227.5	5.8	1486	4 Q9UQ08	Q9UQ08 homo sapien
10	215.5	5.5	1510	5 Q25920	Q25920 plasmodium
11	210.5	5.4	1584	4 Q9UQ09	Q9UQ09 homo sapien
12	195.5	5.0	1661	5 Q06166	Q06166 plasmodium
13	187.5	4.8	2083	5 Q9NA35	Q9NA35 caenorhabdi
14	185.5	4.8	3147	5 Q17464	Q17464 caenorhabdi
15	185.5	4.8	3484	5 P91257	P91257 caenorhabdi
16	181.5	4.7	2081	10 Q9LH98	Q9LH98 arabidopsis

17	179.5	4.6	1790	3	Q07380	Q07380 saccharomyc
18	177	4.5	1211	5	Q23260	Q23260 caenorhabdi
19	176.5	4.5	1265	4	Q9P274	Q9P274 homo sapien
20	176.5	4.5	1323	5	Q9NB35	Q9NB35 plasmodium
21	175	4.5	823	4	Q9H0H6	Q9H0H6 homo sapien
22	174	4.5	725	5	Q965E8	Q965E8 entamoeba h
23	174	4.5	911	3	Q06704	Q06704 saccharomyc
24	174	4.5	2603	6	Q9XS53	Q9XS53 potorous tr
25	174	4.5	17352	5	Q95YM2	Q95YM2 procamburus
26	172.5	4.4	1038	13	Q90784	Q90784 gallus gall
27	171	4.4	1812	5	Q9VZD9	Q9VZD9 drosophila
28	171	4.4	3257	5	Q9V736	Q9V736 drosophila
29	169.5	4.3	2287	5	Q19761	Q19761 caenorhabdi
30	169	4.3	1229	5	Q94185	Q94185 caenorhabdi
31	169	4.3	1616	4	Q96PH3	Q96PH3 homo sapien
32	168.5	4.3	1125	16	Q98PM9	Q98PM9 mycoplasma
33	168.5	4.3	3704	5	P91904	P91904 caenorhabdi
34	168	4.3	594	3	Q9P4A3	Q9P4A3 emericella
35	168	4.3	808	4	O15083	O15083 homo sapien
36	168	4.3	832	4	Q9NUJ4	Q9NUJ4 homo sapien
37	168	4.3	1185	4	Q9NQL1	Q9NQL1 homo sapien
38	168	4.3	1235	4	Q9H2G2	Q9H2G2 homo sapien
39	167.5	4.3	456	11	Q9CS77	Q9CS77 mus musculu
40	167.5	4.3	853	10	Q9LFE4	Q9LFE4 arabidopsis
41	167.5	4.3	1287	5	Q9NGS5	Q9NGS5 dictyosteli
42	167.5	4.3	2363	4	Q99968	Q99968 homo sapien
43	167	4.3	1327	4	Q9Y2L2	Q9Y2L2 homo sapien
44	167	4.3	1395	2	Q9AIS0	Q9AIS0 staphylococ
45	166.5	4.3	927	3	Q9HES9	Q9HES9 emericella

ALIGNMENTS

RESULT 1

O60934	PRELIMINARY;	PRT;	754 AA.
ID	O60934;	O60672;	
AC	O60934;	O60672;	
DT	01-AUG-1998 (TREMBLrel. 07, Created)		
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)		
DE	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	NIBRIN (NIJMEGEN BREAKAGE SYNDROME PROTEIN 1) (CELL CYCLE REGULATORY PROTEIN P95).		
DE	NBS1 OR NBS.		
GN	NBS1 OR NBS.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID:9606;		
RN	[1]		
RX	SEQUENCE FROM N.A., AND VARIANT GLN-553.		
RX	MEDLINE-98250062; PubMed-9590180;		
RA	Varon R., Vissinga C., Platzer M., Cerosaletti K.M., Chrzanowska K.H.,		
RA	Saak K., Beckmann G., Seemanova E., Cooper P.R., Nowak N.J., Stumm M.,		
RA	Weenae C.M.R., Gatti R.A., Wilson R.K., Digweed M., Rosenthal A.,		
RA	Sperling K., Concannon P., Reis A.;		
RT	"Nibrin, a novel DNA double-strand break repair protein, is mutated in		
RT	Nijmegen breakage syndrome.";		
RL	Cell 93:467-476(1998).		
RN	[2]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE-98282099; PubMed-9620777;		
RA	Matsuura S., Tauchi H., Nakamura A., Kondo N., Sakamoto S., Endo S.,		
RA	Smeets D., Solder B., Belohradsky B.H., Kaloustian V.M., Oshimura M.,		
RA	Isonura M., Nakamura Y., Komatsu K.;		
RT	"Positional cloning of the gene for Nijmegen breakage syndrome.";		
RT	Nat. Genet. 19:179-181(1998).		
RL	[3]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE-99134304; PubMed-9933573;		
RA	Tauchi H., Matsuura S., Isonura M., Kinjo T., Nakamura A.,		
RA	Sakamoto S., Kondo N., Endo S., Komatsu K., Nakamura Y.;		
RT	"sequence analysis of an 800-kb genomic DNA region on chromosome 8q21		
RT	that contains the Nijmegen breakage syndrome gene, NBS1.";		

Genomics 55:242-247(1999).
 [4]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=98250063; PubMed=9590181;
 RA Carney J.P., Maser R.S., Olivares H., Davis E.M., Le Beau M.,
 RA Yates J.R. III, Hays L., Morgan W.F., Petrini J.H.J.;
 RT "The hMre11/hRad50 protein complex and Nijmegen breakage syndrome:
 RT linkage of double-strand break repair to the cellular DNA damage
 response.";
 RL Cell 93:477-486(1998).
 CC -1- FUNCTION: INVOLVED IN REPAIR OF DNA DOUBLE-STRAND BREAKS. WORKS
 CC AS PART OF A COMPLEX WHICH ALSO INCLUDES RAD50 AND MRE11.
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS. EXPRESSED AT HIGH LEVELS IN
 CC TESTIS.
 CC -1- DISEASE: DEFECTS IN NBS1 ARE THE CAUSE OF NIJMEGEN BREAKAGE
 CC SYNDROME, AN AUTOSOMAL RECESSIVE CHROMOSOMAL INSTABILITY SYNDROME
 CC CHARACTERIZED BY MICROCEPHALY, GROWTH RETARDATION,
 CC IMMUNODEFICIENCY, CANCER PREDISPOSITION, CELL CYCLE CHECKPOINT
 CC DEFECTS AND IONIZING RADIATION SENSITIVITY.
 DR EMBL: AF051334; AAC39732.1; -;
 DR EMBL: AB013139; BAA28616.1; -;
 DR EMBL: AF049895; AAD08722.1; -;
 DR EMBL: AF069291; AAC62232.1; -;
 DR EMBL: AF058696; AAC39752.1; -;
 DR MIM: 251260; -;
 DR MIM: 602667; -;
 DR InterPro: IPR001357; BRCT.
 DR InterPro: IPR000253; FHA_domain.
 DR Pfam: PF00533; BRCT; 1.
 DR Pfam: PF00498; FHA; 1.
 DR SMART: SM00292; BRCT; 1.
 DR SMART: SM00240; FHA; 1.
 DR PROSITE: PS50006; FHA_DOMAIN; 1.
 KW DNA repair; Glycoprotein; Polymorphism.
 FT VARIANT 185 185 E -> Q.
 FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 142 142 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 440 440 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 600 600 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 666 666 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 754 AA; 84958 MW; CD602F09BA73DAB6 CRC64;
 Query Match 99.9%; Score 3896; DB 4; Length 754;
 Best Local Similarity 99.9%; Pred. No. 1.7e-220;
 Matches 753; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MWKLLPAAGPAGGEPYRLLTGVVYVGRKNCALIIENDQISIRNHAVALTANFVTLNLSQT 60
 DB 1 MWKLLPAAGPAGGEPYRLLTGVVYVGRKNCALIIENDQISIRNHAVALTANFVTLNLSQT 60
 QY 61 DEIPVLTKDNSKYGFVNEEKQNGFSRTLKSGDGTITFGVFGSKFRIEYPLVACSSCL 120
 DB 61 DEIPVLTKDNSKYGFVNEEKQNGFSRTLKSGDGTITFGVFGSKFRIEYPLVACSSCL 120
 QY 121 DVSGKTALNAQILQLGGFTVNNNTEECTHLVMVSVKVTIKTICALICGRPIVAKPEYFTEF 180
 DB 121 DVSGKTALNAQILQLGGFTVNNNTEECTHLVMVSVKVTIKTICALICGRPIVAKPEYFTEF 180
 QY 181 LKAVQSKKQPPIESFYPLDEPISGKNVDLSGRQERKQIFKGTIFILNAKQHKLLSS 240
 DB 181 LKAVESKKQPPIESFYPLDEPISGKNVDLSGRQERKQIFKGTIFILNAKQHKLLSS 240
 QY 241 AVVFGGEARLITEENEEENFFLAPGTCVVDGTGINTNSQTLIPDCQKKWTQSIWMDLQRO 300
 DB 241 AVVFGGEARLITEENEEENFFLAPGTCVVDGTGINTNSQTLIPDCQKKWTQSIWMDLQRO 300
 QY 301 GLRPIPEAETGLAVIFMTTKNYCDPOGHPSTGLKTTTPGSLSQSVSDKLMPSAPVNT 360
 DB 301 GLRPIPEAETGLAVIFMTTKNYCDPOGHPSTGLKTTTPGSLSQSVSDKLMPSAPVNT 360

QY 361 TTYVADTSEQADTWLDSERPKEIKVSKMBQKFRMLSQDAPTVKESCKTSSNNNSMYSNT 420
 DB 361 TTYVADTSEQADTWLDSERPKEIKVSKMBQKFRMLSQDAPTVKESCKTSSNNNSMYSNT 420
 QY 421 LAKMRIPNYQLSPTKLPKINSKDRASQOQQTNSIRNYFOPSTYKRRERDEENQEMSSCKS 480
 DB 421 LAKMRIPNYQLSPTKLPKINSKDRASQOQQTNSIRNYFOPSTYKRRERDEENQEMSSCKS 480
 QY 481 ARLETSCSLLEQQTQPATPSPSLWKNKEQHLSENEPVDTSNDNNLFTDCLKSTVKNASKSH 540
 DB 481 ARLETSCSLLEQQTQPATPSPSLWKNKEQHLSENEPVDTSNDNNLFTDCLKSTVKNASKSH 540
 QY 541 AAEKLRSNKKRMDVAIEDEVLEQLFKDTKPELEIDVKVQKEEDVNVKRPRMDIETN 600
 DB 541 AAEKLRSNKKRMDVAIEDEVLEQLFKDTKPELEIDVKVQKEEDVNVKRPRMDIETN 600
 QY 601 DTFSDEAVPSSKISQENEIGKRELKEDSLWSAKEISNNDKLQDDSEMLPKLLLTFR 660
 DB 601 DTFSDEAVPSSKISQENEIGKRELKEDSLWSAKEISNNDKLQDDSEMLPKLLLTFR 660
 QY 661 SLVIKNSTSRNPGINDDYQLKFNFKFKVTYPGAGKLPHIIGGSDLIAHARKNTELE 720
 DB 661 SLVIKNSTSRNPGINDDYQLKFNFKFKVTYPGAGKLPHIIGGSDLIAHARKNTELE 720
 QY 721 EWLQRQMEVQNHAKESLADDLFRYNPYLKRRR 754
 DB 721 EWLQRQMEVQNHAKESLADDLFRYNPYLKRRR 754
 RESULT 2
 Q9R207 PRELIMINARY; PRT; 751 AA.
 AC Q9R207;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE NIBRIN.
 GN NBN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20108791; PubMed=10640816;
 RA Vissing C.S., Yeo T.C., Woessner J., Massa H.F., Wilson R.K.,
 RA Trask B.J., Concannon P.;
 RT "Identification, characterization, and mapping of a mouse homolog of
 RL Cytoogenet. Cell Genet. 87:80-84(1999).
 DR EMBL: AF076687; AAD20943.1; -;
 DR MGD: MGI:1351625; Nbn.
 DR InterPro: IPR001357; BRCT.
 DR InterPro: IPR000253; FHA_domain.
 DR Pfam: PF00533; BRCT; 1.
 DR Pfam: PF00498; FHA; 1.
 DR SMART: SM00292; BRCT; 1.
 DR SMART: SM00240; FHA; 1.
 DR PROSITE: PS50006; FHA_DOMAIN; 1.
 SQ SEQUENCE 751 AA; 83794 MW; C9F597CC08227B2C CRC64;
 Query Match 69.5%; Score 2710; DB 11; Length 751;
 Best Local Similarity 70.9%; Pred. No. 5.7e-151;
 Matches 537; Conservative 81; Mismatches 129; Indels 10; Gaps 9;
 QY 1 MWKLLPAAGPAGGEPYRLLTGVVYVGRKNCALIIENDQISIRNHAVALTANFVTLNLSQT 60
 DB 1 MWKLLPAAGPAGGEPYRLLTGVVYVGRKNCALIIENDQISIRNHAVALTANFVTLNLSQT 60
 QY 61 DEIPVLTKDNSKYGFVNEEKQNGFSRTLKSGDGTITFGVFGSKFRIEYPLVACSSCL 120
 DB 61 DEIPVLTKDNSKYGFVNEEKQNGFSRTLKSGDGTITFGVFGSKFRIEYPLVACSSCL 120

Db 61 DEIPTLTIKDNGSKYGTGVNEEKMQTGLSCTLKTDGRTVFGVFSKFRVEYEPVLVGCSSCL 120
QY 121 DVSGKTALNOAILQGLGFTVNNWTEECTHLVMSVKTITKICALICGRPIVKPEYFTEF 180
Db 121 DVSGKTALNOAILQGLGFTVNNWTEECTHLVMSVKTITKICALICGRPIVKPEYFSEF 180
QY 181 LKAVQSKKQPPQIESYPPYLDPEPSIGSKNVDSGROKQIFKQKTFIFLNAKQHKLLS 240
Db 181 LKAVESKKQPPDIESYPPYLDPEPSIGSKNVDSGROKQIFKQKTFIFLNAKQHKLLS 240
QY 241 AVVFGGGEARLITEENEEHNFLAPGTCVVDGITSOTLIPDCQKQWJQSTMDMLQRO 300
Db 241 AVVFGGGEARLMAEDDEEESQFSAPGTCVVDGITSOTLIPDCQKQWJQSTMDMLQRO 300
QY 301 GLRPIPEAEIGLAVIFMTTKNYCDPOGHPSTGLKTTTPGPSLQGVSDVKLMPSPVNT 360
Db 301 GLRPIPEAEIGLAVIFMTTKNYCDPOGHPSTGLKTTTPGPSLQGVSDVKLMPSPVNT 360
QY 361 TTYVADTESQADT-WDLSEPKKEIKVSKMEQKFRMLSDQAPTVKESCKTSSNNNSMVSN 419
Db 361 TTYVADTESQADT-WDLSEPKKEIKVSKMEQKFRMLSDQAPTVKESCKTSSNNNSMVSN 419
QY 420 TLAKMRIPNYQSLPTKLPINSKDRASQOQNTSNRNFQPSQKREDEENQEMSSCK 479
Db 420 TLAKMRIPNYQSLPTKLPINSKDRASQOQNTSNRNFQPSQKREDEENQEMSSCK 479
QY 421 TLVRGKTPSYQLSPMKFPVANKNDWTSQOQ-NSIKNYFQPCTRKREDEENPELSCK 479
Db 421 TLVRGKTPSYQLSPMKFPVANKNDWTSQOQ-NSIKNYFQPCTRKREDEENPELSCK 479
QY 480 SARIETSCSLEQTOPATPSLWKNKEQHLSENEPVDNNDNLFTDLDLSIVKNSASKS 539
Db 480 SARIETSCSLEQTOPATPSLWKNKEQHLSENEPVDNNDNLFTDLDLSIVKNSASKS 539
QY 540 HAAEKLRSNKKREMDVVAIEDEVLEQLFKDTPKPELIDVVKVQKOEEDVNVKRPMDIET 599
Db 540 HAAEKLRSNKKREMDVVAIEDEVLEQLFKDTPKPELIDVVKVQKOEEDVNVKRPMDIET 599
QY 599 LPTETLPRKRKDV-DLATEEVEVLEELLRSTKPELAVQVKEQAEADDTIRKKPRMDAER 597
Db 599 LPTETLPRKRKDV-DLATEEVEVLEELLRSTKPELAVQVKEQAEADDTIRKKPRMDAER 597
QY 600 NDTFSDAEVPESSKISQENIEGKKRELKEDSLWSAK-EISNNDKLODDSEMLPKLLLTE 658
Db 600 NDTFSDAEVPESSKISQENIEGKKRELKEDSLWSAK-EISNNDKLODDSEMLPKLLLTE 658
QY 659 FRSILVTK--NSTSRNPSGINDYQGLKNFKKFKKVTYPGAGKLPHIIGGSDLIHAHARKN 716
Db 659 FRSILVTK--NSTSRNPSGINDYQGLKNFKKFKKVTYPGAGKLPHIIGGSDLIHAHARKN 716
QY 671 TELEEWLRQEMEVQONHAKESLADDLFRYNPYLKR 753
Db 671 TELEEWLRQEMEVQONHAKESLADDLFRYNPYLKR 753
QY 715 TELEEWLRQEMEVQONHAKESLADDLFRYNPNVKKR 751
Db 715 TELEEWLRQEMEVQONHAKESLADDLFRYNPNVKKR 751
RESULT 3
ID O88981 PRELIMINARY; PRT; 751 AA.
AC O88981;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE NIBRIN.
GN NBN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN STEM;
RA Mas C., Bourgeois F., Simonneau M.;
RT "Isolation of 50 cDNAs differentially expressed in embryonic forebrain
as compared to mid and hindbrain : a strategy to identify candidate
genes involved in human neurodevelopmental diseases."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF092840; AAC62113.1;
DR MGD; MGI:1351625; Nbn.
DR InterPro; IPR001357; BRCT.
DR InterPro; IPR000253; FHA_domain.
DR Pfam; PF00533; BRCT; 1.

DR Pfam; PF00498; FHA; 1.
DR SMART; SM00292; BRCT; 1.
DR SMART; SM00240; FHA; 1.
DR PROSITE; PS50006; FHA_DOMAIN; 1.
SQ SEQUENCE 751 AA; 83826 MW; 9D136BCC23DC51D9 CRC64;
Query Match 69.4%; Score 2705; DB 11; Length 751;
Best Local Similarity 70.8%; Pred. No. 11e-150;
Matches 536; Conservative 82; Mismatches 129; Indels 10; Gaps 9;
QY 1 MWKLLPAAGPAGGEPYRLLTGVVEYVVGKKNCAILIENDQISRNHAYLTAFVSNLSQT 60
Db 1 MWKLLPAAGPAGGEPYRLLTGVVEYVVGKKNCAILIENDQISRNHAYLTAFVSNLSQT 60
QY 61 DELPVLTLKDNSKYGTGVNEEKMQGFSRSLKSGDGTTFGVGSKFRVEYEPVLVGCSSCL 120
Db 61 DELPVLTLKDNSKYGTGVNEEKMQGFSRSLKSGDGTTFGVGSKFRVEYEPVLVGCSSCL 120
QY 121 DVSGKTALNOAILQGLGFTVNNWTEECTHLVMSVKTITKICALICGRPIVKPEYFTEF 180
Db 121 DVSGKTALNOAILQGLGFTVNNWTEECTHLVMSVKTITKICALICGRPIVKPEYFSEF 180
QY 181 LKAVQSKKQPPQIESYPPYLDPEPSIGSKNVDSGROKQIFKQKTFIFLNAKQHKLLS 240
Db 181 LKAVESKKQPPDIESYPPYLDPEPSIGSKNVDSGROKQIFKQKTFIFLNAKQHKLLS 240
QY 241 AVVFGGGEARLITEENEEHNFLAPGTCVVDGITSOTLIPDCQKQWJQSTMDMLQRO 300
Db 241 AVVFGGGEARLMAEDDEEESQFSAPGTCVVDGITSOTLIPDCQKQWJQSTMDMLQRO 300
QY 301 GLRPIPEAEIGLAVIFMTTKNYCDPOGHPSTGLKTTTPGPSLQGVSDVKLMPSPVNT 360
Db 301 GLRPIPEAEIGLAVIFMTTKNYCDPOGHPSTGLKTTTPGPSLQGVSDVKLMPSPVNT 360
QY 361 TTYVADTESQADT-WDLSEPKKEIKVSKMEQKFRMLSDQAPTVKESCKTSSNNNSMVSN 419
Db 361 TTYVADTESQADT-WDLSEPKKEIKVSKMEQKFRMLSDQAPTVKESCKTSSNNNSMVSN 419
QY 420 TLAKMRIPNYQSLPTKLPINSKDRASQOQNTSNRNFQPSQKREDEENQEMSSCK 479
Db 420 TLAKMRIPNYQSLPTKLPINSKDRASQOQNTSNRNFQPSQKREDEENQEMSSCK 479
QY 480 SARIETSCSLEQTOPATPSLWKNKEQHLSENEPVDNNDNLFTDLDLSIVKNSASKS 539
Db 480 SARIETSCSLEQTOPATPSLWKNKEQHLSENEPVDNNDNLFTDLDLSIVKNSASKS 539
QY 540 HAAEKLRSNKKREMDVVAIEDEVLEQLFKDTPKPELIDVVKVQKOEEDVNVKRPMDIET 599
Db 540 HAAEKLRSNKKREMDVVAIEDEVLEQLFKDTPKPELIDVVKVQKOEEDVNVKRPMDIET 599
QY 599 LPTETLPRKRKDV-DLATEEVEVLEELLRSTKPELAVQVKEQAEADDTIRKKPRMDAER 597
Db 599 LPTETLPRKRKDV-DLATEEVEVLEELLRSTKPELAVQVKEQAEADDTIRKKPRMDAER 597
QY 600 NDTFSDAEVPESSKISQENIEGKKRELKEDSLWSAK-EISNNDKLODDSEMLPKLLLTE 658
Db 600 NDTFSDAEVPESSKISQENIEGKKRELKEDSLWSAK-EISNNDKLODDSEMLPKLLLTE 658
QY 659 FRSILVTK--NSTSRNPSGINDYQGLKNFKKFKKVTYPGAGKLPHIIGGSDLIHAHARKN 716
Db 659 FRSILVTK--NSTSRNPSGINDYQGLKNFKKFKKVTYPGAGKLPHIIGGSDLIHAHARKN 716
QY 671 TELEEWLRQEMEVQONHAKESLADDLFRYNPYLKR 753
Db 671 TELEEWLRQEMEVQONHAKESLADDLFRYNPYLKR 753
QY 715 TELEEWLRQEMEVQONHAKESLADDLFRYNPNVKKR 751
Db 715 TELEEWLRQEMEVQONHAKESLADDLFRYNPNVKKR 751
RESULT 4
Q9R1X1 PRELIMINARY; PRT; 751 AA.
ID Q9R1X1;
AC Q9R1X1;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE NIBRIN.
GN NBN OR NBS1.

OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=MIXTURE OF BRAIN, AND TESTIS;	
RA	Saito T.;	
RT	"Structure of the mouse Nijmegen breakage syndrome (Nibrin/Nbs1) protein."	
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AB016988; BAA76298.1; -.	
DR	MGD; MGI:1351625; Nbn	
DR	InterPro; IPR001357; BRCT.	
DR	InterPro; IPR002533; FHA_Domain.	
DR	Pfam; PF00533; BRCT; 1.	
DR	Pfam; PF00498; FHA; 1.	
DR	SMART; SM00292; BRCT; 1.	
DR	SMART; SM00240; FHA; 1.	
DR	PROSITE; PS50006; FHA_DOMAIN; 1.	
SO	SEQUENCE 751 AA; 83842 MW; ABA75A61F57838F9 CRC64;	
Query Match 69.3%; Score 2702; DB 11; Length 751;		
Best Local Similarity 70.7%; Pred. No. 1.7e-150;		
Matches 535; Conservative 82; Mismatches 130; Indels 10; Gaps 9;		
QY	1 MWKLLPAAGPAGGPPYRLLTGVEYVGRKNCALIIENDQSISRHNHVLTFANFVNTLSQT 60	
DB	1 MWKLLPAAGAAGPYPYRLLAGVEYVGRKNCALIIENDQSISRHNHVLTFVNFVTSLSQT 60	
QY	61 DEIPVLTKDMSKYGTFFVNEEKMONGSRITKSGDGTFFGVGSKFRLEYEPLVACSSCL 120	
DB	61 DEIPVLTKDMSKYGTFFVNEEKMONGSRITKSGDGTFFGVGSKFRLEYEPLVACSSCL 120	
QY	121 DVSGKTALNQAILQGGFTVNNWTEECTHLVMSVKKVTIKTICALICGRPIVKPEYTFEF 180	
DB	121 DVSGKTALNQAILQGGFTVNNWTEECTHLVMSVKKVTIKTICALICGRPIVKPEYTFEF 180	
QY	181 LKAVOSKKQPPQIESFYPPDEPAIGSKSVDSLGRHHERKQIFKCKTFVFLNAKHKLSS 240	
DB	181 LKAVOSKKQPPQIESFYPPDEPAIGSKSVDSLGRHHERKQIFKCKTFVFLNAKHKLSS 240	
QY	241 AVFVGGEARLITENEENHFFLAGTCVVDGTITNSOTLIPDCOKKQIOSIMDLQOR 300	
DB	241 AVFVGGEARLITENEENHFFLAGTCVVDGTITNSOTLIPDCOKKQIOSIMDLQOR 300	
QY	301 GLRPIPAEIGLAVIFMTTKNYCDPQGPSTGLTKTTTPGSLSGVSKFVDEKIMPSAPVNT 360	
DB	301 GLRPIPAEIGLAVIFMTTKNYCDPQGPSTGLTKTTTPGSLSGVSKFVDEKIMPSAPVNT 360	
QY	361 TTYVADTSEQADT-WDLSEPRKEIKVSKMBQKFRMLSQDAPTVKESCKTSSNNNSMVSN 419	
DB	361 TTYVADTSEQADT-WDLSEPRKEIKVSKMBQKFRMLSQDAPTVKESCKTSSNNNSMVSN 419	
QY	420 TLAKMRIPNQLSTKLPSINKSKDRASQOQOTNSIRNYQPSTKKREREENQEMSSCK 479	
DB	420 TLAKMRIPNQLSTKLPSINKSKDRASQOQOTNSIRNYQPSTKKREREENQEMSSCK 479	
QY	480 SARLETSCSLLEQTPATPSLWKNKEOHLSENEPVDNTSDNNLTFTDLSKIVKNSAKS 539	
DB	480 SARLETSCSLLEQTPATPSLWKNKEOHLSENEPVDNTSDNNLTFTDLSKIVKNSAKS 539	
QY	540 HAAEKLKSKKREDDVAIEDEVLQFKDTKPELEIDVKVQKEEDVNVKRRPRMDIET 599	
DB	540 HAAEKLKSKKREDDVAIEDEVLQFKDTKPELEIDVKVQKEEDVNVKRRPRMDIET 599	
QY	599 LPTETLPRKKVDV-DLATEEEVLEELRSTKPELAVQVKEQADDTIRKPRMDAER 597	
DB	599 LPTETLPRKKVDV-DLATEEEVLEELRSTKPELAVQVKEQADDTIRKPRMDAER 597	
QY	600 NDTFSDPAVPESSKISONEIGKRELKEDSLASAK-EISNNDKLQDDSEMLPKLLLTE 658	
DB	600 NDTFSDPAVPESSKISONEIGKRELKEDSLASAK-EISNNDKLQDDSEMLPKLLLTE 658	
QY	658 NRPLNGSEPSALQEDREKELQTES-WSTKHEIANSGLQDSSEELPKLLLTE 716	
DB	658 NRPLNGSEPSALQEDREKELQTES-WSTKHEIANSGLQDSSEELPKLLLTE 716	

DB	657 FRSLVSVNHNSTSRNL-C-VN-EGGLKKNFKKATFFCAGKPLHIIGSDLVGHARKN 714	
QY	717 TELEEWLRQEMEVQNOHAKESLADDLFRYNPNYKRR 753	
DB	715 TELEEWLRQEMEVQNOHAKESLADDLFRYNPNYKRR 751	
RESULT 5		
ID	Q9JIL9 PRELIMINARY; PRT; 750 AA.	
AC	Q9JIL9;	
DT	01-OCT-2000 (TrEMBLrel. 15, Created)	
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)	
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	
DE	NBS1.	
GN	NBS1.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=20368653; PubMed=10908350;	
RA	Lanson N.A. Jr.; Egeland D.B.; Royals B.A.; Claycomb W.C.;	
RT	"The MRE11-NBS1-RAD50 pathway is perturbed in SV40 large T antigen-	
RT	immortalized At-1, At-2 and HL-1 cardiomyocytes.";	
RL	Nucleic Acids Res. 28:2882-2892(2000).	
DR	EMBL; AF218575; AAF91228.1; -.	
DR	InterPro; IPR001357; BRCT.	
DR	InterPro; IPR002533; FHA_Domain.	
DR	Pfam; PF00533; BRCT; 1.	
DR	Pfam; PF00498; FHA; 1.	
DR	SMART; SM00292; BRCT; 1.	
DR	SMART; SM00240; FHA; 1.	
DR	PROSITE; PS50006; FHA_DOMAIN; 1.	
SO	SEQUENCE 750 AA; 83304 MW; 7368BAD0914CF305 CRC64;	
Query Match 68.9%; Score 2685.5; DB 11; Length 750;		
Best Local Similarity 69.5%; Pred. No. 1.6e-149;		
Matches 525; Conservative 93; Mismatches 130; Indels 7; Gaps 7;		
QY	1 MWKLLPAAGPAGGPPYRLLTGVEYVGRKNCALIIENDQSISRHNHVLTFANFVNTLSQT 60	
DB	1 MWKLLPAASAAPGPPYRLLAGVEYVGRKNCALIIENDQSISRHNHVLTFVNFVTSLSQT 60	
QY	61 DEIPVLTKDMSKYGTFFVNEEKMONGSRITKSGDGTFFGVGSKFRLEYEPLVACSSCL 120	
DB	61 DEIPVLTKDMSKYGTFFVNEEKMONGSRITKSGDGTFFGVGSKFRLEYEPLVACSSCL 120	
QY	121 DVSGKTALNQAILQGGFTVNNWTEECTHLVMSVKKVTIKTICALICGRPIVKPEYTFEF 180	
DB	121 DVSGKTALNQAILQGGFTVNNWTEECTHLVMSVKKVTIKTICALICGRPIVKPEYTFEF 180	
QY	181 LKAVOSKKQPPQIESFYPPDEPAIGSKSVDSLGRHHERKQIFKCKTFVFLNAKHKLSS 240	
DB	181 LKAVOSKKQPPQIESFYPPDEPAIGSKSVDSLGRHHERKQIFKCKTFVFLNAKHKLSS 240	
QY	241 AVFVGGEARLITENEENHFFLAGTCVVDGTITNSOTLIPDCOKKQIOSIMDLQOR 300	
DB	241 AVFVGGEARLITENEENHFFLAGTCVVDGTITNSOTLIPDCOKKQIOSIMDLQOR 300	
QY	301 GLRPIPAEIGLAVIFMTTKNYCDPQGPSTGLTKTTTPGSLSGVSKFVDEKIMPSAPVNT 360	
DB	301 GLRPIPAEIGLAVIFMTTKNYCDPQGPSTGLTKTTTPGSLSGVSKFVDEKIMPSAPVNT 360	
QY	361 TTYVADTSEQADT-WDLSEPRKEIKVSKMBQKFRMLSQDAPTVKESCKTSSNNNSMVSN 419	
DB	361 TTYVADTSEQADT-WDLSEPRKEIKVSKMBQKFRMLSQDAPTVKESCKTSSNNNSMVSN 419	
QY	420 TLAKMRIPNQLSTKLPSINKSKDRASQOQOTNSIRNYQPSTKKREREENQEMSSCK 479	
DB	420 TLAKMRIPNQLSTKLPSINKSKDRASQOQOTNSIRNYQPSTKKREREENQEMSSCK 479	
QY	421 TLVRGKAPNQLSPMKCPAAKKNKDWSS-QOQLNSIRNYFQPSKKREREENQEQSSCK 479	
DB	421 TLVRGKAPNQLSPMKCPAAKKNKDWSS-QOQLNSIRNYFQPSKKREREENQEQSSCK 479	

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Metkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J.C., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Wellenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C., "The genome sequence of Drosophila melanogaster";
 RT Science 287:2185-2195(2000).
 RL EMBL; AE003551; AAF50215.1; -.
 DR FlyBase; FBgn0026198; nbs.
 DR InterPro; IPR001357; BRCT.
 DR InterPro; IPR002523; FHA_domain.
 DR Pfam; PF00498; FHA; 1.
 DR SMART; SM00292; BRCT; 1.
 DR SMART; SM00240; FHA; 1.
 DR PROSITE; PS00006; FHA_DOMAIN; 1.
 DR KW Hypothetical protein.
 SQ SEQUENCE 811 AA; 90286 MW; 2BB9936468B05E8E CRC64;

Query Match 12.1%; Score 473; DB 5; Length 811;
 Best Local Similarity 23.8%; Pred. No. 8.3e-20;
 Matches 199; Conservative 129; Mismatches 297; Indels 210; Gaps 31;

Qy 8 AGPAGGEPYRLLTGVE-VVVGKKNCAILLNDQSTSRNHAVLTANFVSNLSQTDIEPVL 66
 Db 17 AAIADEKFLVLPFGKVTIGRLATDLIVAQDLISRNHAQLIQ-----TEADGGDTL 70
 Qy 67 TLKD-NSKYGTGFV---NEEKMQGFSRT---LKSGDGTTFGFGSKFRIEYPLVACSSC 119
 Db 71 HIEDLSRYGTFIFPKNSQKPRKVPKSTPLPVGTRLRFGANMSIWQVTKLVITYSA 130
 Qy 120 LDVSGKTALNAIQLGGFTVNNWTECTHLVMVSVKVTIKTICALICGRPIVPEYFTE 179
 Db 131 LTRSEVQLTKMLPEPMGGTGTSTNTEEGSHLTMTNEVSVTVKLLHMLNKPITVTPPYWK 190
 Qy 180 FLKAVQS---KKPOPIESFYPLDEPSIGSKNVDLSGRQERKQIFKGTIFLNAKQHK 236
 Db 191 MLQAAQSILHVKEGPPQEDYQ-----TNIDVTWRPRTLFAKGTIVEMNRKHFED 241
 Qy 237 KLSAVVFGGGEARLITENEENHFFLAPGTCVVDITGITSQTLIPQCCKKWTQSIMDM 296
 Db 242 MYGWSVQKAGATCDI---NSGVRKTFLTKSDVIV-----IQVPSQSQATSEINNI 291
 Qy 297 LQROGLRPIPAETGLAVIFMTTKNCDPOGHP--STGLKTTTPGPSLQSGVSDKELMP 354
 Db 292 LEQNGRRIIQIEYIGMALIHCSITEFCNPT-HKFTSDSLPT-----ESVTS 337
 Qy 355 SAPVNTTYYVADTE--SQADTWLSEK--PKEIKVSKWEQKFMLSODAPTVKE--SKK 408
 Db 338 SMARNSSIIVPTEHSAQSNAATPLSELVVPSECEMEQDASQDSQAKRRSHAS 397
 Qy 409 TSSNNNSVSNLAK-----MRIPN-----YQLSPTKLPS-----INKSK 443
 Db 398 TVDSSDEKKTSLSKRAKSDIATKLTMTKSKNAILLDSSLEEDVTIPAPAPVQVTRQSK 457

Qy 444 DRASQQOQNTSIRNFQPSQTKKRER-----DEENQEMSMCKSARIETSCSLLEQTOPA 496
 Db 458 AIAEKEVHPVPVPAASKHITRTKQVFCVDSDEENENARKPK-----ETPAPT 506
 Qy 497 TPSLWKKEQHLSENEPVDNTSDNNLFTDTDLKSLVKKSASKSAAEKLRSNKKREMDVV 556
 Db 507 IPSMAKKKTE-----APVATRISPRL-NGKSLATNITNPADKHAVP-----AKRPVLVS 555
 Qy 557 AIEDEVLE-QLFKDK-K-PELEIDVKVQ----- 581
 Db 556 ASSDEEDGDLFQPKSPQKPAETVVQPRIAGKNAPARISVVDFLEKSQAQEPAPVPQ 615
 Qy 582 -KQEDVNVRRKRMIDETNDFTSDEAVP-----ESSKISQENEIGCKRELKEDSLWSAKE 636
 Db 616 LESQSQTPRKRLRLEL-LNESDSDCDNLFNFADSKKKRKTQEAQRNDDSDTGLFNFS 674
 Qy 637 ISNNDKLQDDSEML-----PKLLLLTFERSLIVKNSISRP- 672
 Db 675 ERPSDHDDEDSRLTEPFPVPETESKKQSKYIVAPRRDRPKKVDISDADSVMKETSIAKDPD 734
 Qy 673 -----SGINDD---YGOLKNFKFKKVTYP 694
 Db 735 EEQWLAAMKDSIEVRMCNLNIVIRSQEEVDASLEDSVKNHGGKKNKFKFVKTKNP 789
 RESULT 8
 ID 075058 PRELIMINARY; PRT; 1460 AA.
 AC 075058;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE KIAA0470 PROTEIN.
 GN KIAA0470 OR KAB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=98115662; PubMed=9455484;
 RA Seki N., Ohira N., Nagase T., Ishikawa K., Miyajima N., Nakajima D.,
 RA Nomura N., Ohara O.;
 RT "Characterization of cDNA clones in size-fractionated cDNA libraries
 from human brain.";
 RL DNA Res. 4:345-349(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Hara Y., Adachi Y.;
 RT "Molecular cloning and initial characterization of KAB.";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB007939; BAA32315.1; -;
 DR EMBL; AB022659; BAA83380.1; -;
 DR InterPro; IPR000253; FHA_domain.
 DR Pfam; PF00498; FHA; 1.
 DR SMART; SM00240; FHA; 1.
 DR PROSITE; PS00006; FHA_DOMAIN; 1.
 SQ SEQUENCE 1460 AA; 161436 MW; BAD23EBCA19B65F0 CRC64;

Query Match 5.8%; Score 227.5; DB 4; Length 1460;
 Best Local Similarity 18.8%; Pred. No. 4.2e-05;
 Matches 154; Conservative 149; Mismatches 318; Indels 199; Gaps 29;

Qy 2 WKLLPAAGPAGEPYRLLTGVEYVVVGKKNCAILLNDQSTSRNHAVLTANFVSNLSQTD 61
 Db 6 WFLVSSGGTRHLRPREMI-----FVGRODCELMQLQS-RSVDKQHAVINYDAS-----TD 53
 Qy 62 EIPVLTLKDNKSYGTFVNEEKMQGFSRTLKSGDGTITGFGSKFRIEYPLVACSSCLD 121
 Db 54 EHLVKDL--GSLNGTFVNDVRIPEQTYITLKLKDLRFEGYDTNLTFTV----- 98


```
QY 662 LVIKNSTSRNPSGI-----NDYGO-----LKNFKFKKVTYPGAGKLPHIIGGSDLIAH 711
: ||| | | : | : ||| |
Db 1303 ATWNLNIIIFASGIATREENTDVLVEERIQRVEFKTT-----1343

QY 712 HARKNTELE-EWLQRQEMEVQNHAKESLA-----DDLFF 744
: | | : | : | | | | : | : | | |
Db 1344 ---ENLEIQKEVVLTKEEVDNSDVKEHRTSAVNIDLDVDF 1380

RESULT 15
P91257
ID P91257 PRELIMINARY; PRT: 3484 AA.
AC P91257;
DT 01-MAY-1997 (TremBLrel. 03, Created)
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE HYPOTHETICAL 385.7 KDA PROTEIN.
GN F12F3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Fulton B., Wohlmann P.;
RT "The sequence of C. elegans cosmid F12F3.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U80022; AAC25885.2; -.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR003962; FnIII_repeat.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_Like.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00041; fn3; 7.
DR Pfam; PF00047; ig; 10.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SM00060; FN3; 8.
DR SMART; SM00409; IG; 12.
DR SMART; SM00408; Igc2; 8.
DR SMART; SM00410; IG_Like; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Hypothetical protein; Immunoglobulin domain; Repeat.
SQ SEQUENCE 3484 AA; 385746 MW; 21DD6F6893020F4F CRC64;

Query Match 4.8%; Score 185.5; DB 5; Length 3484;
Best Local Similarity 20.0%; Pred. No. 0.035;
Matches 132; Conservative 114; Mismatches 237; Indels 177; Gaps 31;

QY 166 IGRPIVKPEYTFELKAVQSKK---OPQIESFYPLDPEPSIGSKNVDLSGRQERKQIF 222
: | | : | | : | | : | | : | | : | |
Db 466 VTGKPL-----EAKKPVEDKKDASQSSSKESPPTD-----KKQIP 505
```

```
QY 223 KGKTFIPL-----NAQHKHKLSSAVVFGGE-----ARLITEEN 256
: | | : | | : | | : | | : | |
Db 506 KA---LFIPDEISSRFCDPSTMHSETNITTTIRREGSADAKTLPVEPLSASVMKVESA 562

QY 257 EEEHNFFLAPGTCVVDGTINSOTLIPDCQKWKIQSMDMLQROGLRPIPEARTIGLAVIF 316
: | : | | | : | | : | | : | | : | | : | |
Db 563 KEKAEFSFK-----RRSET--PDDKSR-----KKEGLPPAKKSE-----594

QY 317 MITKNYCDPQGHPSGTGLKTTTPGSPSLSQGVSDVDE-KILMPSAP-----VNTTIVYADT 367
: | | : | | : | | : | | : | |
Db 595 -----KKDEVTAEKQSTEALESKKKEVDESKISEQOPSDKNKSEVGVPERAAGP 645

QY 368 ESEQADTWLSERPKEIKVSKMEQKF-RMLSQDAPTVKESKTSSTNNNSMVSWTAKMRI 426
: | : | : | | : | | : | | : | | : | | : | |
Db 646 ETKK-DVSEIEEVPKPKTKIKKTEKSDSSISQKSNVLKPADDDKSKSDDDVTDKSKKTTT- 703

QY 427 PNQLSPTKLPSINKSKDRASQOQOTNSIRNYFPQSTPKKRERDEEENQEMSSCKSARLET 486
: | : | : | : | : | : | : | : | : | : | : |
Db 704 -----DOTKVATDSKLEKAADTTKQIET-ETVVDDKSKKVKLKKKTEKSDSFISQKSETP 757

QY 487 CSLEQOTQATPSLWKNKEQHLSENEFPVDTNNDNNLFTDLDKSIYKNKSASKSHAAEKL 546
: | | : | | : | | : | | : | | : | | : | |
Db 758 -PVVEPTKPAESEAQKIAEVNKAQKQ---KEVDNLKREAEV-----AAKKIADKELK 806

QY 547 -----SNKKREMD--VAIEDEVLEQLFKDTKPELEIDVKVQKQEEEDVNR 590
: | : | : | : | : | : | | : | | : | | : | |
Db 807 IEAEANIKKTAEEVAAAKQEKDEQLKLETEVYVSKSAAEKLELEKQAQIKKAAEADAVK 866

QY 591 KRPRMDIETND-----TFSDAEVPRESSKISOEN-EIGKKRELK-----ED 629
: | : | : | : | : | : | : | : | : | : | : |
Db 867 KQKELN-EKNKLEAAKSAADKLKLEESAAKSKKSESVKFGEEKKTAGEKTVQVES 925

QY 630 SLWSAKEISNND-KLQDDSEMLPKLLLTFRSLVKNSTSRNPNGIND----DYGOLKN 684
: | | : | : | : | : | | : | | : | | : | |
Db 926 EPTSKKTTIDTKDVGATEPADETPKK-----KIIKKTEKSDSSISQKSDATDSEKVS 977

QY 685 FKFKKVTYPGAGKLPHIIGGSDLIAHHARKNTELEWLRQEMEV---QNHAKESLAD 741
: | : | | : | | : | : | : | : | : | : | : |
Db 978 QKEQDEPTKPAVSETQMVTE-----ADKSKKQKETDEKLEKLDIAEIAAKTKQEADEKSKLD 1032
```

Search completed: August 15, 2002, 01:48:48
Job time: 488 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:32:45 ; Search time 438.62 Seconds
(without alignments)
17234.906 Million cell updates/sec

Title: US-09-837-602-1

Perfect score: 4403

Sequence: 1 ttgcgcacgagcgcggttg.....accgcggtggagctccagct 4403

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4105.2	93.2	4386	21	AAZ89047
2	4103.6	93.2	4386	21	AAZ34997
3	1862.8	42.3	6521	21	AAZ89046
4	1836.4	41.7	2044	22	AAH13813
5	671	15.2	752	22	AAH03660
C 6	552.2	12.5	646	21	AAH78056
C 7	552.2	12.5	646	22	AAI28794
C 8	520.4	11.8	543	22	AAH09188
9	358.4	8.1	396	22	AAF44923

10	193	4.4	195	16	AAT24560	Human gene signatu
11	163.4	3.7	6440	22	AAK68762	Human immune/haema
12	163.4	3.7	6440	22	AAK68763	Human immune/haema
13	163.4	3.7	6440	22	AAK68764	Human immune/haema
14	120.2	2.7	121	16	AAT25563	Human gene signatu
15	119.4	2.7	442	22	AAK06689	Human brain expres
16	119.4	2.7	442	22	AAK32388	Human bone marrow
17	119.4	2.7	442	22	AAI38231	Probe #6917 used t
18	114	2.6	114	22	AAI19606	Human brain expres
19	114	2.6	114	22	AAK45612	Human bone marrow
20	114	2.6	114	22	AAI51546	Probe #20232 used
21	97.6	2.2	6237	24	ABL32358	Human immune syste
C 22	92.4	2.1	6237	24	ABL32359	Human immune syste
23	72.2	1.6	7442	22	AAK46686	Tumour suppressor
24	72	1.6	14006	24	ABL33958	Chemically pretrea
25	71	1.6	14006	24	ABL33958	Human immune syste
C 26	70	1.6	6056	22	AAK46363	Tumour suppressor
C 27	70	1.6	6056	24	ABL32780	Human immune syste
28	69.6	1.6	15548	22	ABL34155	Human immune syste
C 29	69.2	1.6	2787	22	AAO7987	Human ovarian and
C 30	69.2	1.6	2787	22	AAI05661	Human reproductive
C 31	69.2	1.6	2787	22	AAK69792	Human immune/haema
C 32	68.8	1.6	7442	22	AAK46686	Tumour suppressor
C 33	68.2	1.5	6171	24	ABL32788	Human immune syste
34	68.2	1.5	6668	24	ABL33697	Human immune syste
35	68.2	1.5	8392	24	ABL33490	Human immune syste
C 36	68	1.5	6644	20	AAK33181	Base sequence of t
C 37	68	1.5	7372	20	AAK33182	Base sequence of t
C 38	68	1.5	7797	20	AAK33180	Cowpox virus bsr f
C 39	68	1.5	7796	20	AAK33184	Base sequence of t
C 40	68	1.5	68940	20	AAK57351	Human chromosome 6
41	67.2	1.5	6292	22	AAK46735	Tumour suppressor
42	67	1.5	6145	24	ABL32972	Human immune syste
43	67	1.5	6644	20	AAK33181	Base sequence of t
44	67	1.5	7372	20	AAK33182	Base sequence of t
45	67	1.5	7797	20	AAK33180	Cowpox virus bsr f

ALIGNMENTS

RESULT 1

AAZ89047

ID AAZ89047 standard; DNA; 4386 BP.

XX AAZ89047;

AC AAZ89047;

XX 01-JUN-2000 (first entry)

DT Human nibrin DNA.

DE Nibrin; human; DNA double strand break repair protein; diagnosis;

KW therapy; Nijmegen Breakage Syndrome; gene therapy; ds.

XX Homo sapiens.

OS Key

FT CDS

FT Location/Qualifiers

FT 26..2290

FT /*tag= a

FT /product= "Nibrin"

XX DE19818680-C1.

XX 09-MAR-2000.

XX 27-APR-1998; 98DE-1018680.

XX 27-APR-1998; 98DE-1018680.

XX (UYBE) UNIV BERLIN HUMBOLDT.

XX WPI; 2000-196117/18.

XX P-PSDB; AAY51669.

DR

XX A DNA double strand break repair protein, Nibrin, and related DNA
 PT useful for diagnosis and therapy of Nijmegen Breakage Syndrome and
 PT other diseases influenced by DNA-double-strand break repair -
 XX
 XX Claim 2a; Fig 2; 32pp; German.
 XX
 CC This invention describes a novel DNA double strand break repair protein,
 CC Nibrin. Nibrin and DNA encoding it are useful for diagnosis and/or
 CC therapy of diseases influenced by repair of DNA-double strand breaks,
 CC in particular Nijmegen Breakage Syndrome. The product of the invention
 CC has applications in gene therapy. This sequence encodes the nibrin
 CC protein described in the invention.
 XX
 SQ Sequence 4386 BP; 1467 A; 744 C; 845 G; 1330 T; 0 other;

Query Match 93.28; Score 4105.2; DB 21; Length 4386;
 Best Local Similarity 97.1%; Pred. No. 0;
 Matches 4270; Conservative 7; Mismatches 30; Indels 91; Gaps 6;
 QY 28 gccccagccctgagagccgacccgatgtgaaaactctgcccgccgcccgcgcgagc 87
 DB 1 gccccagccctgagagccgacccgatgtgaaaactctgcccgccgcccgcgcgagc 60
 QY 88 agggagaaccatacacagacttttgaactggtgagtagctgttggaaaggaaactgtgc 147
 DB 61 agggagaaccatacacagacttttgaactggtgagtagctgttggaaaggaaactgtgc 120
 QY 148 cattcattgaaatgatacagtcagtcagccgaataatcgtctgttaactgttaactt 207
 DB 121 cattcattgaaatgatacagtcagtcagccgaataatcgtctgttaactgttaactt 180
 QY 208 ttctgttaaccaactgagtcagtcagtcagccgaataatcgtctgttaactgttaactt 267
 DB 181 ttctgttaaccaactgagtcagtcagtcagccgaataatcgtctgttaactgttaactt 240
 QY 268 taagtgtgtaccttgttgaatgaggaaaaaatacagaatggctttcccgaaactttgaa 327
 DB 241 taagtgtgtaccttgttgaatgaggaaaaaatacagaatggctttcccgaaactttgaa 300
 QY 328 gtcggggagatgtattacttttgagtggtttggaagtaaatccagaatagatagacc 387
 DB 301 gtcggggagatgtattacttttgagtggtttggaagtaaatccagaatagatagacc 360
 QY 388 ttgtgtgcatgctctcttctgttagatgctctggaactgcttlaaatcaagctat 447
 DB 361 ttgtgtgcatgctctcttctgttagatgctctggaactgcttlaaatcaagctat 420
 QY 448 attgcaacttggagatttactgttaacaaattggacagaagaatgcactcacttgtcat 507
 DB 421 attgcaacttggagatttactgttaacaaattggacagaagaatgcactcacttgtcat 480
 QY 508 ggtatcagtgaaagtaccattaaacaaatattgtgcactcatttggagcgtcccaattgt 567
 DB 481 ggtatcagtgaaagtaccattaaacaaatattgtgcactcatttggagcgtcccaattgt 540
 QY 568 aaagccagaatatttactgaattccctgaaagcaggttcagtcaccaagcagcgtccaca 627
 DB 541 aaagccagaatatttactgaattccctgaaagcaggttcagtcaccaagcagcgtccaca 600
 QY 628 aattgaaagtgtttaccacacttcttgatgaacctctatttggaaagtataattgtgtatct 687
 DB 601 aattgaaagtgtttaccacacttcttgatgaacctctatttggaaagtataattgtgtatct 660
 QY 688 gtcaggagccgagaaagaaacaaatcttcaaaagggaacatttatatttttgaatgc 747
 DB 661 gtcaggagccgagaaagaaacaaatcttcaaaagggaacatttatatttttgaatgc 720
 QY 748 caaacagcataagaaattgattgtccgcagttgtctcttggaggtgggaagcagtagttgat 807
 DB 721 caaacagcataagaaattgattgtccgcagttgtctcttggaggtgggaagcagtagttgat 780

QY 808 aacagaagagaatgaagaagaacataatttttttggctccgggaacgtgtgttgtga 867
 DB 781 aacagaagagaatgaagaagaacataatttttttggctccgggaacgtgtgttgtga 840
 QY 868 tacagggaataaacaactcacagacaccttaattcctgactgtcagaagaatgattcagtc 927
 DB 841 tacagggaataaacaactcacagacaccttaattcctgactgtcagaagaatgattcagtc 900
 QY 928 aataatgatatgtccaaaaggcgaaggtcttagacctattcctgaagcagaataattggatt 987
 DB 901 aataatgatatgtccaaaaggcgaaggtcttagacctattcctgaagcagaataattggatt 960
 QY 988 ggcggtgatttcatgactacaaagaattactgtgactcctcagggccatccacgtacagc 1047
 DB 961 ggcggtgatttcatgactacaaagaattactgtgactcctcagggccatccacgtacagc 1020
 QY 1048 attaaagacaaacaactccaggaccaaagcctttccaaaggcgtgtcagttgtgataaaact 1107
 DB 1021 attaaagacaaacaactccaggaccaaagcctttccaaaggcgtgtcagttgtgataaaact 1080
 QY 1108 aatgccaaagcgcacagtgaaacactacaacatacgtagctgacacagatcagagcaagc 1167
 DB 1081 aatgccaaagcgcacagtgaaacactacaacatacgtagctgacacagatcagagcaagc 1140
 QY 1168 agatcacatgggatttggatgaaaggccaaagaatacaaaagctctccaaatggaaacaaa 1227
 DB 1141 agatcacatgggatttggatgaaaggccaaagaatacaaaagctctccaaatggaaacaaa 1200
 QY 1228 attcagaatgctttcacaagacgcccactgtaaaagggtcctgcacaaacagctctaa 1287
 DB 1201 attcagaatgctttcacaagacgcccactgtaaaagggtcctgcacaaacagctctaa 1260
 QY 1288 taataatagtagtgatacaataactttggctaaagtagagaatcccaactcagcttctc 1347
 DB 1261 taataatagtagtgatacaataactttggctaaagtagagaatcccaactcagcttctc 1320
 QY 1348 accaactaaattgccaagtagtaataaaagtaaaagtagggtctctcagcagcagcagac 1407
 DB 1321 accaactaaattgccaagtagtaataaaagtaaaagtagggtctctcagcagcagcagac 1380
 QY 1408 caactccatcagaagaacttacttccagcgtctaccacaaaggaaggaaggaagaaa 1467
 DB 1381 caactccatcagaagaacttacttccagcgtctaccacaaaggaaggaaggaagaaa 1440
 QY 1468 tcaagaatgctcttcagtaaatcagcaagaatgaaacgtctgttctctttttagaaca 1527
 DB 1441 tcaagaatgctcttcagtaaatcagcaagaatgaaacgtctgttctctttttagaaca 1500
 QY 1528 aacacacactgtctacacccctcattgttggaataataagagcagcatctatctgagaatga 1587
 DB 1501 aacacacactgtctacacccctcattgttggaataataagagcagcatctatctgagaatga 1560
 QY 1588 gctgtggacacaaactcagacaaataacttattacagatacagatttataaactattgt 1647
 DB 1561 gctgtggacacaaactcagacaaataacttattacagatacagatttataaactattgt 1620
 QY 1648 gaaaaattctgcccagtaaatctcattgctgcagaaagtaagatacaataaaaaaaggga 1707
 DB 1621 gaaaaattctgcccagtaaatctcattgctgcagaaagtaagatacaataaaaaaaggga 1680
 QY 1708 aatggatgtagtggccatagaagatgaagtattggaacgtatttcaaggacacaaaacc 1767
 DB 1681 aatggatgtagtggccatagaagatgaagtattggaacgtatttcaaggacacaaaacc 1740
 QY 1768 agaattagaattgtagtgaaagttcaaaacaggaagaagatgtcaattgttagaaaaag 1827
 DB 1741 agaattagaattgtagtgaaagttcaaaacaggaagaagatgtcaattgttagaaaaag 1800
 QY 1828 gccaaagatgtagatagaacaaatgacactttcagtgatgaagcagtcaccagaagtag 1887
 DB 1801 gccaaagatgtagatagaacaaatgacactttcagtgatgaagcagtcaccagaagtag 1860
 QY 1888 caaaatatctcaagaataatgaaattgggaagaaacgtgaactcaaggaagcactcattgt 1947

Db	1861		caaaa	tatctc	aagaaa	tgaaat	tgga	gaagaa	crgaa	ctca	agga	gac	ccactatg	1920			
Qy	1948	gtcag	ctaa	agaaa	tatctc	aaacat	gacaa	actc	agga	gtatag	ttagat	gtctcc	aaaa	2007			
Db	1921	gtcag	ctaa	agaaa	tatctc	aaacat	gacaa	actc	agga	gtatag	ttagat	gtctcc	aaaa	1980			
Qy	2008	aaag	ctgt	tatt	gact	gaaatt	tagat	ca	ctggt	gatt	taaaa	actct	acttcc	agaatcc	2067		
Db	1981	aaag	ctgt	tatt	gact	gaaatt	tagat	ca	ctggt	gatt	taaaa	actct	acttcc	agaatcc	2040		
Qy	2068	gtctg	gcata	aaat	gatg	att	gtgt	cca	actaaaa	attt	caagaaa	tcca	aaaggt	ctac	2127		
Db	2041	atctg	gcata	aaat	gatg	att	gtgt	cca	actaaaa	attt	caagaaa	tcca	aaaggt	ctac	2100		
Qy	2128	ata	ctctg	gag	cagg	aaaa	acttcc	acac	tatc	at	tgagg	atc	agatcta	atag	ctcatca	2187	
Db	2101	ata	ctctg	gag	cagg	aaaa	acttcc	acac	tatc	at	tgagg	atc	agatcta	atag	ctcatca	2160	
Qy	2188	tgctg	aa	aga	aata	caga	aactaga	agag	tggct	ag	cagg	aaaa	tggag	gtac	aaaa	2247	
Db	2161	tgctg	aa	aga	aata	caga	aactaga	agag	tggct	ag	cagg	aaaa	tggag	gtac	aaaa	2220	
Qy	2248	aca	tgca	aa	aga	agag	gtctct	gtgt	g	at	gtat	gata	caaat	co	cttatt	2307	
Db	2221	aca	tgca	aa	aga	agag	gtctct	gtgt	g	at	gtat	gata	caaat	co	cttatt	2280	
Qy	2308	gaga	ga	ta	aact	gag	gattt	taaaa	ga	gcac	at	ggaaa	acttcc	ctag	taag	ctcatc	2367
Db	2281	gaga	ga	ta	aact	gag	gattt	taaaa	ga	gcac	at	ggaaa	acttcc	ctag	taag	ctcatc	2340
Qy	2368	ttcag	gcc	aa	ca	ag	gtt	tata	tga	tata	tat	gtat	gata	ga	gcgtatt	taag	2427
Db	2341	ttcag	gcc	aa	ca	ag	gtt	tata	tga	tata	tat	gtat	gata	ga	gcgtatt	taag	2400
Qy	2428	tta	tg	cc	ta	aatt	tata	taaa	ta	aatgc	acaa	actt	gatt	ctt	gtt	gtat	2487
Db	2401	tta	tg	cc	ta	aatt	tata	taaa	ta	aatgc	acaa	actt	gatt	ctt	gtt	gtat	2460
Qy	2488	tg	tt	gt	ct	gt	ttc	ag	gctt	gtc	at	gt	cat	ctt	ttt	ctatt	2547
Db	2461	tg	tt	gt	ct	gt	ttc	ag	gctt	gtc	at	gt	cat	ctt	ttt	ctatt	2520
Qy	2548	tt	t	ta	a	a	ta	g	t	a	t	a	g	t	c	a	2607
Db	2521	tt	t	ta	a	ta	g	t	a	g	t	c	a	g	t	c	2579
Qy	2608	ta	a	g	t	c	a	c	c	t	t	c	a	c	a	t	2640
Db	2580	ta	a	g	t	c	a	c	c	t	t	c	a	c	a	t	2639
Qy	2641	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2648	
Db	2640	t	a	c	t	a	t	c	t	a	g	a	a	t	t	t	2699
Qy	2649	a	a	t	a	t	t	c	a	c	a	a	a	t	a	t	2708
Db	2700	a	a	t	a	t	t	c	a	c	a	a	a	t	a	t	2759
Qy	2709	c	t	a	g	t	a	a	a	a	t	a	g	a	a	t	2768
Db	2760	c	t	a	g	t	a	a	a	a	t	a	g	a	a	t	2818
Qy	2769	c	a	t	a	g	t	a	a	a	t	a	g	a	a	t	2828
Db	2819	c	a	t	a	g	t	a	a	a	t	a	g	a	a	t	2878
Qy	2829	g	a	t	c	a	c	c	a	t	a	g	t	a	a	a	2888
Db	2879	g	a	t	c	a	c	c	a	t	a	g	t	a	a	a	2938
Qy	2889																

Db	2939	acatggggttagtatttcaacagaatgaaagacctgtaccccttttgggtggtcttactg	2998
Qy	2949	agctggggtgggtgctgtgttttgagcttatttagagtcctagtttctactataaagta	3008
Db	2999	agctggggtgggtgctgtgttttgagcttatttagagtcctagtttctactataaagta	3058
Qy	3009	gaatgggtgagatgtgttttttttttctaccktaagaggagatggtagaacaatgaatg	3068
Db	3059	gaatgggtgagatgtgttttttttttctaccktaagaggagatggtagaacaatgaatg	3118
Qy	3069	tccttttcaaacattattgacaagtgaatttccaagtctgtgttcaaaaatatattcatg	3128
Db	3119	tccttttcaaacattattgacaagtgaatttccaagtctgtgttcaaaaatatattcatg	3178
Qy	3129	tacctgtgatccagcaagaaggagttccagtcgaagagtcactacaactgattagttggt	3188
Db	3179	tacctgtgatccagcaagaaggagttccagtcgaagagtcactacaactgattagttggt	3238
Qy	3189	tagaagaatgagaataatgaaacagtgaggaatggaggccaatatctccatgaacttccactt	3248
Db	3239	tagaagaatgagaataatgaaacagtgaggaatggaggccaatatctccatgaacttccactt	3298
Qy	3249	aacagaagaacaagaagggaacaagagctggcctctacatcaactctcaccttccaaatct	3308
Db	3299	aacagaagaacaagaagggaacaagagctggcctctacatcaactctcaccttccaaatct	3358
Qy	3309	tgtggaagtgcattctacttgcacagaaccaaataacttacttccaagtctctggctgctg	3368
Db	3359	tgtggaagtgcattctacttgcacagaaccaaataacttacttccaagtctctggctgctg	3418
Qy	3369	caggtggaaactccagctgcgaaggagttagggaataagaagtccttttttaaaagcttct	3428
Db	3419	caggtggaaactccagctgcgaaggagttagggaataagaagtccttttttaaaagcttct	3478
Qy	3429	cagccttctcagggaacaagaataatgggtgagccaactctgcaatttctactacaggaactg	3488
Db	3479	cagccttctcagggaacaagaataatgggtgagccaactctgcaatttctactacaggaactg	3538
Qy	3489	agaccagttagatatttgaataatatagagagttatgaaacacttaaatatgatgagtg	3548
Db	3539	agaccagttagatatttgaataatatagagagttatgaaacacttaaatatgatgagtg	3598
Qy	3549	tatgacattggatagaacaatggatcttttagaagtagaattgcagggccatatattagctg	3608
Db	3599	tatgacattggatagaacaatggatcttttagaagtagaattgcagggccatatattagctg	3658
Qy	3609	atgaaatgagtcatttgagtcctttaaagccaatgatacataattaccaagtgaaactg	3668
Db	3659	atgaaatgagtcatttgagtcctttaaagccaatgatacataattaccaagtgaaactg	3718
Qy	3669	gtggaacatagtgtctccattttacagttgaagggaataaaaggacagattaaattattgty	3728
Db	3719	gtggaacatagtgtctccattttacagttgaagggaataaaaggacagattaaattattgty	3778
Qy	3729	tctgtcatgcccacaataccctttctaagggaagactgcccctactatagcagtttttatatt	3788
Db	3779	tctgtcatgcccacaataccctttctaagggaagactgcccctactatagcagtttttatatt	3838
Qy	3789	tgtcaatttatgaataataatgaatgaggagttctgtgacctctgtctttacaacaatttg	3848
Db	3839	tgtcaatttatgaataataatgaatgaggagttctgtgacctctgtctttacaacaatttg	3897
Qy	3849	gggtgtgccagttattttcccttttttaaccmcttcccaattccgaattcggtgtgtagtgatg	3908
Db	3898	g-----tgtgtcagttattttcccttttttaacccattccaatcggtgtgtagtgatg-----	3948
Qy	3909	ttccaatttgggttttaatttggatatccctgtagtagcataaattgggtccatagaaaattctt	3968
Db	3949	tttcaatttgggttttaatttggatatccctgtagtagcataaattgggtccatagaaaattctt	4008
Qy	3969	tatacattcttagatgcaagtcctctgycgggataacgtattgagataattacacctagctct	4028
Db	4009	tatacattcttagatgcaagtcctctgycgggataacgtattgagataattacacctagctct	4068

QY 73 cgcgggcccgcaggagagagaaaccatacacagacttttgactgctgcttgagtagcgtttgttgg 132
Db 127 cgcgggcccgcaggagagagaaaccatacacagacttttgactgctgcttgagtagcgtttgttgg 186
QY 133 aaggaaaaactgtccattcttaattgaaatgatcagtcgcatcagcagcgaataatcgtgt 192
Db 187 aaggaaaaactgtccattcttaattgaaatgatcagtcgcatcagcagcgaataatcgtgt 246
QY 193 gtaactgctaacttttctgtaaccacactgagtcacaaacagatgaaatccctgtattgac 252
Db 247 gtaactgctaacttttctgtaaccacactgagtcacaaacagatgaaatccctgtattgac 306
QY 253 attaaaagataattctaagtattgtaaccttctgtaagtgaagaaaaatgcagaatggctt 312
Db 307 attaaaagataattctaagtattgtaaccttctgtaagtgaagaaaaatgcagaatggctt 366
QY 313 ttccggaacttgaagtcgggggatgtattacttttggagtggttggaaataattcag 372
Db 367 ttccggaacttgaagtcgggggatgtattacttttggagtggttggaaataattcag 426
QY 373 aatagagtatgacgtttgtgcatgctcttctgttttagatgctctgggaaactgc 432
Db 427 aatagagtatgacgtttgtgcatgctcttctgttttagatgctctgggaaactgc 486
QY 433 tttaaatcaagctatattgcaacttggaggatttactgttaacaaattggacagaagaatg 492
Db 487 tttaaatcaagctatattgcaacttggaggatttactgttaacaaattggacagaagaatg 546
QY 493 cactcacccttgcattgatacagtgaaagtaccattaaacaaatatgtgcactcatttg 552
Db 547 cactcacccttgcattgatacagtgaaagtaccattaaacaaatatgtgcactcatttg 606
QY 553 tggacgtccaattgttaaagccagaaatattttactgaaattcctgaaagcagttcagtccaa 612
Db 607 tggacgtccaattgttaaagccagaaatattttactgaaattcctgaaagcagttcagtccaa 666
QY 613 gaagcagcctccacaaattgaaagtttttaccacaccttcttgatgaaccatctattgg-aa 671
Db 667 gaagcagcctncacaaattgaaagtttttaccacaccttcttgatgaaccatctattggaa 726
QY 672 gtaaaaatgttgatctgcaggacgg 697
Db 727 gtaaaaatgttgatctgcaggacgg 752

RESULT 6
AAA78056/c
ID AAA78056 standard; cDNA; 646 BP.
XX
AC AAA78056;
XX
DT 14-NOV-2000 (first entry)
XX
DE cDNA encoding human colon tumour polypeptide, SEQ ID NO:343.
XX
KW Human colon tumour polypeptide; tumour antigen; cancer; vaccine;
KW immunotherapy; diagnosis; progression; ss.
XX
OS Homo sapiens.
XX
PN WO200037643-A2.
XX
PD 29-JUN-2000.
XX
PF 23-DEC-1999; 99WO-US50909.
XX
PR 23-DEC-1998; 98US-0221298.
PR 02-JUL-1999; 99US-0347496.
PR 22-SEP-1999; 99US-0401064.
PR 19-NOV-1999; 99US-0444242.
PR 02-DEC-1999; 99US-0454150.

(CORI-) CORIXA CORP.

Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk J;
Wang T, Yuqiu J;

WPI; 2000-442671/38.

New colon tumor polypeptides used to inhibit the development of cancer,
especially colon cancer, and for diagnosing and monitoring the
progression of the cancer -

Claim 29; Page 190; 229pp; English.

Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or
portions of proteins which are associated with human colon tumours.
The invention also specifically discloses 8 human colon tumour proteins
(AAB11897-B11904). The nucleic acids, the polypeptides they encode, and
antigen presenting cells (APCs), preferably dendritic cells) expressing
such polypeptides may be used in vaccines that target tumour cells,
especially colon tumour cells, thereby inhibiting the development of
cancer. T-cells specific for the polypeptide expressed by the APC are
used to remove tumour cells from biological samples, especially blood or
fractions thereof. The sample or the isolated T-cells specific for the
polypeptide can then be used to inhibit cancer development. CD4+ and/or
CD8+ T-cells from a patient may be incubated with a polypeptide or
nucleic acid of the invention, or an APC expressing such a polypeptide,
to cause the proliferation of specific T-cells. The T-cells can be
cloned and then administered back to the patient to inhibit cancer
development. Nucleic acids encoding the polypeptides and antibodies
against the polypeptides may be used to determine the expression level
of a tumour protein of the invention, and therefore to determine whether
cancer cells are present. Such diagnostic methods may also be used to
monitor the progression of a cancer by repeating the processes at time
intervals, and comparing the current result to previous results. The
present sequence represents a cDNA encoding a human colon tumour
polypeptide.

Sequence 646 BP; 239 A; 124 C; 95 G; 188 T; 0 other;

Query Match 12.5%; Score 552.2; DB 21; Length 646;
Best Local Similarity 93.0%; Pred. No. 1.4e-104;
Matches 609; Conservative 4; Mismatches 32; Indels 10; Gaps 3;

QY 3419 aaaagcttctcagccttctcagggaacagaaatgggtgagccaatctgcaatttctact 3478
Db 645 AAAGCTTCTCAGCCTTCTCAGGGAACAGAAATTTGGGGAGCCCAATCTGCAATTTCTACT 586
QY 3479 acagccattgagaccagtagattattgaaatatatagagagttatgaacactaaatt 3538
Db 585 ACAGGCATTGAGACAGTAGATTATTGAAATATTATAGAGAGTTATGAACACTTAAATT 526
QY 3539 atgatatgtgtagacattgacattgagaaacattcttagaagtagaattgacagggc 3598
Db 525 ATGATAGTGGTATGACATTGGATAGACATGGGATACCTTTTAAAGTAGAATTGACAGGGC 466
QY 3599 atattagttgataaaggagtcattttgagtcctttaaagccatgtatcataataacca 3658
Db 465 ATATTAGTTGATGAATGGAGTCATTGGAGTCTCTTAATAGCCCATGTATCATTAATACCA 406
QY 3659 agtgaagctggggaacatatgttctccattttacagttaaaggaaataataatggacagatt 3718
Db 405 AGTGAAGCTGGTGAACATATATGCTCCATTTTACAGTTAAGGAATATAATTAATGGACAGATT 346
QY 3719 aatattgttctgtcatgcccaaatcccttcttaaggaagactgccctactatagcag 3778
Db 345 AATATTGTCTCTGTCTGTCATGCCCAATCCCTTTCTTAAGGAAGACTGCCCTACTATAGCAG 286
QY 3779 tttttatatttgcataatttatgaataataatgaatgaggttctgggtaccttccctgtcttt 3838
Db 285 TTTTATATTTGTCAATTAATTAATGAATAATAATGAATGA -GAGTTCCTGGTACCTCCCTGTCCTT 227
QY 3839 acaaatattgggtgtgttgcagtatatttccctttttaacccmcttcccaattcgggtgtgt 3898

Db	226	ACAAATATGG-----TGTGTGCAGTATTTTCTTTTAAACCATTCCCAATCGGTGTGTAG	171
QY	3899	aggtggaatgttccattcattgggttttaattgtatatccctgatagctataattgggtcat	3958
Db	170	TGATG-----TTTCATATTTGGTTTTAAATTTGTATATCCCTGATAGCTATAATTGGGTCA	116
QY	3959	agaaattctttatacatcttagatgcgaactctcttgcggatatacgattagaatatta	4018
Db	115	AGAAATTCCTTTATACATCTTAGATGCAATTCCTTGTGCGATATATGTATTGAGATATTA	56
QY	4019	cacctagctctgttgctgactgttttcttttatgtctcttttgatgaatagaagtttt	4073
Db	55	CACCTAGTCGTGGCTGACTGTGTTTCTTTATGTCTCTTTTGTATGATAGAGAAGTTTT	1

RESULT 7
AAI28794/C
ID AAI28794 standard; cDNA: 646 BP.
XX
XX AAI28794;
XX AC
XX
XX 12-OCT-2001 (first entry)
XX DT
XX DE
XX Colon tumour related determined cDNA sequence for clone 25908.

The present invention describes colon tumour associated proteins (I) and the polynucleotides (II) that encode them. (I) have cytostatic activity. (I) and (II) can be used in gene therapy and vaccine production. (I) and (II) may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate colon tumour associated protein (TCAP) expression, such as colonic cancer. For example, (I) and (II) may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of TCAPs by expressing inactive proteins or to supplement the patients own production of them. Additionally, (II) may be used to produce the TCAP proteins, by inserting the nucleic acids into a host cell culturing the cell to express the protein. (II) and its complementary sequences may also be used as DNA probes in diagnostic polymerase chain reaction (PCR) and hybridisation assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. (I) may

CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.
XX
SQ Sequence 646 BP; 239 A; 124 C; 95 G; 188 T; 0 other;

RESULT	8
AAH09188/c	
ID	AAH09188 standard; cDNA; 543 BP.
XX	
XX	
XX	AAH09188;
XX	
XX	
XX	26-JUN-2001 (first entry)
XX	
XX	
DE	Human cDNA clone (3'-primer) SEQ ID NO:6023.
XX	
XX	
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX	

```
OS Homo sapiens.
XX EP1074617-A2.
XX
XX
XX PD 07-FEB-2001.
XX
XX
XX PF 28-JUL-2000; 2000EP-0116126.
XX
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX
XX PA (HELI-) HELIX RES INST.
XX
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI: 2001-318749/34.
XX
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX PS Claim 3; SEQ ID 6023; 2537pp + CD ROM; English.
XX
XX CC The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX SQ Sequence 543 BP; 128 A; 111 C; 83 G; 219 T; 2 other;

Query Match 11.8%; Score 520.4; DB 22; Length 543;
Best Local Similarity 98.3%; Pred. No. 5e-98;
Matches 535; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 1564 ggacagcatctatcagaatgagcctgtggacacaaactcagacaataacttatttac 1623
DB 543 GNAGCAGCTTCTATCTGAGATGAGCCTG-GGACCCAAACTCAGCCCAATAACTTATTTC 485
QY 1624 agatacagattaaactctattgtgaaaaattctgccagtaaatctcatgctgcagaaaa 1683
DB 484 AGATCCAGATTTAAATCTATTGTGAAAAATCTGCCAGTAAATCTCATGCTGCAGAAAA 425
QY 1684 gctaagatcaataaaaaaggggaaatggatgattgtggccatagagaagtattgga 1743
DB 424 GCTAAGATCAATNAAAAAGGAAATGGATGATGTGCCATAGAAAGTAAATATTGGA 365
QY 1744 acagtattcaaggacacaaacacagattagaaattgatgtgaaattcaaaacagga 1803
DB 364 ACAGTTATTCAAGGACACAAANCCAGAGTTAGTGAATTTGATGTGAAGTTCAAAACACGGA 305

QY 1804 ggaagatgtcaatgttagaaaaagcccaagatggtatagaagaaacaaatgacactttcag 1863
DB 304 GGAAGATGTCAATGTTAGAAAAAGGCCAAGGATGGATATAGAAACAATGACACTTTTCAG 245
QY 1864 tgatgaagcagttaccagaaaagtagcaaaaatatctcaagaagaaatgaaattgggaagaaacg 1923
DB 244 TGATGAAGCAGTACCAGAAAGTAGCAAAATATCTCAAGAAAATGAAATTTGGGAAGAAACG 185
QY 1924 tgaactcaaggaagactcactatgtcagctaaagaaatctcaacaatgacaaacttca 1983
DB 184 TGAACCTCAAGGAAGACTCACATGTGTGCTAGCTTAAAGAAATATCTTACAAATGGCAAACTTCA 125
QY 1984 ggatgatgtgagatgcttccaaaaaagctgttattgactgaatttagatcacctgggtgat 2043
DB 124 GGATGATAGTCAGATGCTTCCAAAAAGCTGTTATTGACTGAATTTAGATCAGCTGGTGAT 65
QY 2044 taaaaactctactccagaaaatcgcgtctggcataaaatgatattggtcaactaaaaaa 2103
DB 64 TAAAAACTCTACTTCCAGAAATCCGCTGTGGCATAAATGATGATTATGTCACACTAAAAAA 5
QY 2104 ttcc 2107
DB 4 TTTC 1

RESULT 9
AAF44923
ID AAF44923 standard; cDNA; 396 BP.
XX
XX AC AAF44923;
XX
XX DT 28-MAR-2001 (first entry)
XX
XX DE Human breast cancer related protein coding sequence SEQ ID NO: 79.
XX
XX KW Human; breast cancer; diagnosis; therapy; vaccine; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200078960-A2.
XX
XX PD 28-DEC-2000.
XX
XX XX 23-JUN-2000; 2000WO-US17536.
XX
XX PF 23-JUN-1999; 99US-0140903.
XX
XX PR 12-OCT-1999; 99US-0158980.
XX
XX PA (CORI-) CORIXA CORP.
XX
XX PI Yuqiu J, Mitcham JL;
XX
XX DR WPI: 2001-041426/05.
XX
XX PT New polynucleotides encoding breast tumor specific proteins, useful for
XX prevention, treatment and diagnosis of breast cancer -
XX
XX PS Claim 25; Page 146; 165pp; English.
XX
XX CC The present invention provides the coding sequences for a number of
XX breast cancer related proteins. These can be used in vaccinations
XX against, diagnosis of and treatment of cancer, particularly breast
XX cancer.
XX
XX SQ Sequence 396 BP; 124 A; 52 C; 45 G; 160 T; 15 other;

Query Match 8.1%; Score 358.4; DB 22; Length 396;
Best Local Similarity 93.1%; Pred. No. 1.3e-64;
Matches 353; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

QY 3978 tagatgcaagtctcttgycgatatagctattgagatatcacacttagtctgtggttga 4037
```

Db 1 tanatgcaagtnttngtgatatacgtatttgagataattacnccctagctgtgcttga 60
QY 4038 ctattttcttattgttttgatagaatagagttttaaatttttgacaaagtcgaatttat 4097
Db 61 ctgtttcttcttctgtcttttgatgaatanaagttttaaattttgacaaagtcgaatttat 120
QY 4098 tttttcttttggttgatatttttctctccaatttaaccccgaagatttcagatattctg 4157
Db 121 tttttcttttggttgatatttttctccaatttaaccccgaagatttcanaatattng 180
QY 4158 ctctattataaactttatatattttatttctgtatctaccttgaaattgatattgtatgt 4217
Db 181 ctctattanaaaactttattttattttatttctgtatctaccttgaaattgatattgtatgt 240
QY 4218 tttgaattatggatcaggggtcttttttcccccatacagaattccagtcattgtaacac 4277
Db 241 tttgaattatggatcaggggtcttttttcccccatacagaattccagtcattgtaacac 300
QY 4278 tgtttattgaagaattatcttctctcattaaattacattgccaattagtaaaaatca 4337
Db 301 tgtttattgaanaattatcttctctcattaaattnccttgccaattagtaaaaatca 360
QY 4338 attaacacatrmarmmmrrr 4356
Db 361 attaccataaaaaaaa 379

RESULT 10
AAT24560
ID AAT24560 standard; cDNA to mRNA; 195 BP.
XX
AC AAT24560;
XX
DT 25-SEP-1996 (first entry)
XX
DE Human gene signature HUMGS06610.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX Homo sapiens.
XX
PN WO9514772-A1.
XX
PD 01-JUN-1995.
XX
PF 11-NOV-1994; 94WO-JP01916.
XX
PR 12-NOV-1993; 93JP-0355504.
XX
XX (MATS/) MATSUBARA K.
XX (OKUB/) OKUBO K.
XX Matsubara K, Okubo K;
PI
PA
PA
XX
XX WPI; 1995-206931/27.
XX
XX
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
XX Claim 1; Page 1639; 2245pp; Japanese.
XX
XX A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in AAT19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
SQ Sequence 195 BP; 80 A; 24 C; 31 G; 60 T; 0 other;

Query Match 4.4%; Score 193; DB 16; Length 195;
Best Local Similarity 100.0%; Pred. No. 1.5e-30;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2276 gatctttttgatacaatccttattttaaaggagagataaactgagatttttaaaaga 2335
Db 1 gatctttttgatacaatccttattttaaaggagagataaactgagatttttaaaaga 60
QY 2336 agccatggaaaaacttctcttagtaagcatctacttcaggcccaacaagggttatgaatata 2395
Db 61 agccatggaaaaacttctcttagtaagcatctacttcaggcccaacaagggttatgaatata 120
QY 2396 tagtgtatagaagcgtatttaagttacaatgtttttatgacctaaatttattataaaatg 2455
Db 121 tagtgtatagaagcgtatttaagttacaatgtttttatgacctaaatttattataaaatg 180
QY 2456 cacaaaaactttga 2468
Db 181 cacaaaaactttga 193

RESULT 11
AAK68762
ID AAK68762 standard; DNA; 6440 BP.
XX
AC AAK68762;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23574.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225271.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-02559678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

Disclosure; SEQ ID NO 23574; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I) by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/hematopoietic-related diseases, especially cancers and cancer metastases of hematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/hematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAK82169

RESULT 14
AAT25563
ID AAT25563 standard; cDNA to mRNA; 121 BP.
XX AC AAT25563;
XX XX
DT 15-NOV-1996 (first entry)
XX XX
DE Human gene signature HUMGS07740.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX OS Homo sapiens.
XX PN W09514772-A1.
XX PD 01-JUN-1995.
XX PF 11-NOV-1994; 94WO-JP01916.
XX PR 12-NOV-1993; 93JP-0355504.
XX PA (MATS/) MATSUBARA K.
PA PA (OKUB/) OKUBO K.
XX PI Matsubara K, Okubo K;
XX DR WPI; 1995-206931/27.
XX
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
PS Claim 1; Page 1859; 2245pp; Japanese.
XX
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in AAT19001-726837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
SQ Sequence 121 BP; 41 A; 25 C; 12 G; 43 T; 0 other;

Query Match 2.7%; Score 120.2; DB 16; Length 121;
Best Local Similarity 98.3%; Pred. No. 1.4e-15;
Matches 119; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4229 gatcagggtctttttttcccccatcaagaatccagtcatttaaacactgtttattgaa 4288
Db 1 gatcagggtctttttttcccccatcaagaatccagtcatttaaacactgtttattgaa 60

QY 4289 agaattatcttcttcctattaaattacctgtgccaattagtaaaaaataaattaccatrm 4348
Db 61 agaattatcttcttcctattaaattacctgtgccaattagtaaaaaataaattaccatrm 120

QY 4349 a 4349
Db 121 a 121

RESULT 15
AAK06689
ID AAK06689 standard; DNA; 442 BP.
XX AC AAK06689;
XX XX
DT 05-NOV-2001 (first entry)
XX XX
DE Human brain expressed single exon probe SEQ ID NO: 6680.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX OS Homo sapiens.
XX PN W0200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 6680; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 442 BP; 153 A; 67 C; 80 G; 142 T; 0 other;

Query Match 2.7%; Score 119.4; DB 22; Length 442;
Best Local Similarity 99.2%; Pred. No. 2.9e-15;
Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2118 aaaaagtcacatattcctggagcaggaataacttccacacatcattggaggatcagatctaa 2177
Db 264 aataggtcacatattcctggagcaggaataacttccacacatcattggaggatcagatctaa 323

QY 2178 tagctcatcattgctcgaagaataacacagaaactagaagtgctgaagcaggaaatggagg 2237
Db 324 tagctcatcattgctcgaagaataacacagaaactagaagtgctgaagcaggaaatggagg 383

QY 2238 t 2238
Db 384 t 384

Search completed: August 15, 2002, 00:05:32
Job time: 9167 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 14, 2002, 23:56:50 ; Search time 58.32 Seconds
(without alignments)
1436.036 Million cell updates/sec

Title: US-09-837-602-2
Perfect score: 3899
Sequence: 1 MWKLLPAAGPAGGEPYRLLT.....KEESLADLFYRNPYLKRRR 754

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Length	ID	Description
1	3893	99.8	754	21	AAV51669 Human nibrin prote
2	3885	99.6	754	21	AAV32373 Nibrin, encoded by
3	2046	52.5	399	22	ABR92565 Human protein sequ
4	473	12.1	811	22	ABR62661 Drosophila melanog
5	205	5.3	38	22	AAW60530 Human bone expres
6	205	5.3	38	22	AAW73186 Human protein SEQ
7	205	5.3	38	22	AAW33396 Peptide #7433 enco
8	171.5	4.4	990	22	AAW78520 Human protein SEQ
9	171	4.4	1812	22	ABR58022 Drosophila melanog
10	171	4.4	3257	22	ABR67502 Drosophila melanog
11	169.5	4.3	2918	22	ABG27218 Novel human diagno

12	169	4.3	1128	19	AAW40539 Mutant C-beta prot
13	169	4.3	1128	21	AAV84461 Amino acid sequenc
14	168	4.3	808	22	ABG05140 Novel human diagno
15	168	4.3	832	22	ABR93691 Human protein sequ
16	167.5	4.3	931	22	AAW79504 Human protein SEQ
17	167.5	4.3	2400	22	ABG20278 Novel human diagno
18	167.5	4.3	2415	22	ABG20279 Novel human diagno
19	167	4.3	789	22	AAW5460 Human protein sequ
20	165	4.2	1411	17	AAW02258 Nucleolar/endosoma
21	165	4.2	1886	19	AAW54241 Rattus norvegicus
22	164.5	4.2	1392	20	AAV06999 Restin protein seq
23	164.5	4.2	2703	22	ABR60074 Drosophila melanog
24	163.5	4.2	2139	22	ABR47278 PNT7711. Homo sapi
25	163.5	4.2	3111	22	ABR60327 Drosophila melanog
26	163.5	4.2	6815	22	ABR66811 Drosophila melanog
27	163	4.2	1164	19	AAW40541 Mutant C-beta prot
28	162.5	4.2	1427	12	AAV10534 Human 160KD mediat
29	161	4.1	1093	19	AAW40540 Mutant C-beta prot
30	161	4.1	1129	21	AAW44462 Amino acid sequenc
31	160	4.1	1099	19	AAW40538 Mutant C-beta prot
32	159.5	4.1	1833	22	ABR71141 Drosophila melanog
33	159	4.1	2096	21	AAV41592 Human ORFX ORF1356
34	158.5	4.1	1664	20	AAW9462 C. albicans alpha-1
35	158.5	4.1	2663	22	AAW39097 Human polypeptide
36	158	4.1	726	22	ABR65327 Drosophila melanog
37	158	4.1	1780	22	AAW38681 Human polypeptide
38	158	4.1	2897	22	ABR58514 Drosophila melanog
39	158	4.1	3263	22	ABR67210 Drosophila melanog
40	157.5	4.0	976	22	ABR66581 Human SCP-1 mutain
41	157.5	4.0	1552	22	ABR71764 Drosophila melanog
42	157	4.0	1164	17	AAW85781 Group B Streptococ
43	157	4.0	1164	19	AAW40537 Group B streptococ
44	157	4.0	1164	21	AAW84459 Amino acid sequenc
45	157	4.0	2665	22	ABR28314 Human peptide #965

ALIGNMENTS

RESULT 1
AAV51669
ID AAV51669 standard; Protein; 754 AA.
XX
AC AAV51669;
XX
DT 01-JUN-2000 (first entry)
XX
DE Human nibrin protein.
XX
KW Nibrin; human; DNA double strand break repair protein; diagnosis;
KW therapy; Nijmegen Breakage Syndrome; gene therapy.
XX
OS Homo sapiens.
XX
PN DE19818680-Cl.
XX
PD 09-MAR-2000.
XX
PF 27-APR-1998; 98DE-1018680.
XX
PR 27-APR-1998; 98DE-1018680.
XX
(UYBE) UNIV BERLIN HUMBOLDT.
XX
WPI; 2000-196117/18.
DR N-PSDB; AA289047.
XX
A DNA double strand break repair protein, Nibrin, and related DNA
useful for diagnosis and therapy of Nijmegen Breakage Syndrome and
other diseases influenced by DNA-double-strand break repair -
Clalm 1; Fig 2; 32pp; German.
XX

CC This invention describes a novel DNA double strand break repair protein,
CC Nibrin. Nibrin and DNA encoding it are useful for diagnosis and/or
CC therapy of diseases influenced by repair of DNA-double strand breaks,
CC in particular Nijmegen Breakage Syndrome. The product of the invention
CC has applications in gene therapy. This sequence represents the nibrin
CC protein described in the invention.
XX
SQ Sequence 754 AA;

Query Match 99.8%; Score 3893; DB 21; Length 754;
Best Local Similarity 99.8%; Pred. No. 9.5e-290;
Matches 752; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MWKLLPAAGPAGGPPYRLLTGVEYVVGKNCALIIENDQSIERNHVLITANFSVTNLSQT 60
Db 1 mwklilpaagppgpyrllltgveyvvgkncailiendqsiernhavlitanfsvtnlsq 60

Qy 61 DEIPVLTAKNSKYGTFTVNEEKMONGSRTLKSGDGTFTGFGSKFRIEYPLVACSSCL 120
Db 61 deipvltaknskygtftvneekmngsrtlksgdgtftgfgskfrieypvlvacsscl 120

Qy 121 DVSGKTALNQAILQLGGFTVNNWTEECTHLVMVSVKVTIKTICALICGRPIVKPEYTFE 180
Db 121 dvsgktalnqailqlggftvnnwteecthlvmvsvkvtkticalicgrpivkpeytfef 180

Qy 181 LKAVOSKKQPPQIESFYPPDEPSIGSKNVDSLGRQERKQTFKGTFTFLNAKQHKLS 240
Db 181 lkaveskkppqiesfyppidepsigsknvdslgrqerkqtfkgtftflnakqhkklss 240

Qy 241 AVFEGGEARLITENEENHFFLAPGTCVVDTGITSQTLIPDCQKKWQSIIMDMLOQ 300
Db 241 avfeggearliteeneehfflapgtcvtvdtgitsqtlipdcqkqkqslmdmlq 300

Qy 301 GLRPIEAETGLAVIETTKNYCDQHPSTGLKTTTPGSPLSOGVSDEKLMPAPVNT 360
Db 301 glrpieaeiglaviettknycdqhpstglktttppsgpslsgvsvdekmpsapvnt 360

Qy 361 TTYVADTESEQADTWLSEPKIEKVKMEQKFRMLSQDAPTVMKESCKTSNNNSMVSNT 420
Db 361 ttyvadteseqadtwlserpkelkvsqmeqkfrmlsqdaptvmkescktsnnnsmsvnt 420

Qy 421 LAKRIPNYQLSPKLSINKSKDRASQOQTNSIRNYFPSTKKRERDENQEMSSCKS 480
Db 421 lakrripnyqlspklslksinkskdrasqoqtnsirnyfpstkkrrerdenqemsscks 480

Qy 481 ARIETCSLLEOTOPATPSLWKNKEQHLSENEPVDTSNNLFTDCLKSTVKNASKSH 540
Db 481 arietcsllleotopatpslwknkeqhlsenepvdtssnnlftdclkstivknasasksh 540

Qy 541 AAEKLRNKKREMDVVAIEDEVLEQLFKDTKPELEIDVKQOEDVNVKRPRMDIETN 600
Db 541 aaeeklrnkkremdvvaiedevleqlfkdtkpeleidvkqoedvnnvrkrprmdietn 600

Qy 601 DTFSDEAVPSSKISQENIGCKRELKEDSLWSAKELSNNDKLQDDSEMLPKLLLTFR 660
Db 601 dtfsdeavpsskisoenigckrelkedslwsakelsnndklqddsemplkllltfr 660

Qy 661 SLVTKNFTSRNPSGINDYQOLKFKFKVYTPGACKLPHIIGSDDLIAHARKNTELE 720
Db 661 slvtnkftsrnpsgindyqolkfkfkvytpgacklphiiigsddliaharkntele 720

Qy 721 EWLrqemevqnhakeesladdlfrnpylkr 754
Db 721 ewlrqemevqnhakeesladdlfrnpylkr 754

RESULT 2
ID AAY32373
XX
AC AAY32373;
XX

DT 28-FEB-2000 (first entry)
XX Nibrin, encoded by Nijmegen breakage syndrome gene NBS1.
DE
XX NBS1 gene; nibrin; Nijmegen breakage syndrome; diagnosis; human;
KW gene therapy; cancer; microcephaly; mental retardation;
KW primary ovarian failure.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 24..100
FT /note= "forkhead associated domain"
FT 109..195
FT Domain
FT /note= "breast cancer carboxy-terminal domain"
XX
PN W09955716-A1.
XX
PD 04-NOV-1999.
XX
XX 27-APR-1999; 99WO-US09036.
PF
XX 27-APR-1998; 98US-0083269.
PR
XX (VIRG-) VIRGINIA MASON RES CENT.
PA
XX Concannon PJ, Vissinga CS, Cerosaletti KM, Varon R, Sperling K;
PI Reis A;
PI
PI
PI
DR WPI; 2000-062015/05.
DR N-PSDB; AAZ34997.
XX
XX Novel gene useful for detecting mutations or polymorphisms, and
PT diagnosing certain pathological conditions in Nijmegen Breakage
PT syndrome patients.
PT
PS Claim 1; Fig 2A-B; 58pp; English.
XX
XX This sequence represents nibrin, the protein encoded by the NBS1
CC gene (see AAZ34997) that is associated with the Nijmegen breakage
CC syndrome (NBS). The invention is based on the discovery that the
CC NBS1 gene contains mutations in all NBS patients. These mutations
CC include deletions and insertions that result in frameshift, as well
CC as point mutations. It is an object of the invention to detect a
CC mutation or polymorphism in NBS patients, and hence to diagnose a
CC predisposition to a pathological condition such as cancer,
CC microcephaly, mental retardation, and primary ovarian failure,
CC based on detection of a mutation in the NBS1 gene. It is also an
CC object of the invention to treat NBS by replacing the mutated gene
CC in a NBS patient by gene therapy. Recombinant vectors, genetically
CC engineered host cells, a method for producing nibrin polypeptide,
CC an antibody that specifically binds to the polypeptide, and a
CC method for diagnosing NBS by detection decreased expression of the
CC gene or by detecting a mutation in the gene are claimed.
XX
SQ Sequence 754 AA;

Query Match 99.6%; Score 3885; DB 21; Length 754;
Best Local Similarity 99.6%; Pred. No. 3.9e-289;
Matches 751; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MWKLLPAAGPAGGPPYRLLTGVEYVVGKNCALIIENDQSIERNHVLITANFSVTNLSQT 60
Db 1 mwklilpaagppgpyrllltgveyvvgkncailiendqsiernhavlitanfsvtnlsq 60

Qy 61 DEIPVLTAKNSKYGTFTVNEEKMONGSRTLKSGDGTFTGFGSKFRIEYPLVACSSCL 120
Db 61 deipvltaknskygtftvneekmngsrtlksgdgtftgfgskfrieypvlvacsscl 120

Qy 121 DVSGKTALNQAILQLGGFTVNNWTEECTHLVMVSVKVTIKTICALICGRPIVKPEYTFE 180
Db 121 dvsgktalnqailqlggftvnnwteecthlvmvsvkvtkticalicgrpivkpeytfef 180

QY 181 LKAVQSKKOPQPTESFYPLDPEPSIGSKNVDSLGRQERKQIFKGKTFIFLNAKOHKKLSS 240
Db 181 lkaveskkppqiesfyppldepsigsknvdlsgqrqerqifkgkctfiflnakghkklss 240
QY 241 AVVFGGGEARLITEENEENHFFLAGPTCVVDGTGINSQTLIPDCOKKWTQSIQMDMLQRQ 300
Db 241 avvfgggearliteeneenehnfflapgtcvvdgtginsqtlipdcqckkwigsimdmqlrq 300
QY 301 GLRPIPEAEIGLAVIFMTTKNYCDPGHPSTGLTKTTPGPSLSQGVSVDEKLMPSAPVNT 360
Db 301 glrpipeaeiglavifmttknycdpgghpstgltkttppgslsqgvsvdeklmpsapvnt 360
QY 361 TTYVADTESQADTWDLSERPKEIKVSKMEQKFRMLSDOAPTYSKESCTSSNNNSMVSNT 420
Db 361 ttyvadteseqadtwdlserpkeikvskmeqkfrmlsqdaptvkescktsnnnsmvsnt 420
QY 421 LAKMRIPNYQLSTKPLPSINKSKDRASQOQOTNSIRNYFQPTSKRERDEENQEMSCKS 480
Db 421 lakmripnyqlsptklpsinkskdrasqqgqtnsirnyfqpstkkrerdeengemascks 480
QY 481 ARTECSLLEQTPATPSLWKNKEOHLSENPVDTNSDNNLFTDIDLKSIKNSAKSKSH 540
Db 481 arietcsllseqtpatpslwknkednlsepevdtnsdnnlftdldlksivknsascksh 540
QY 541 AAELRSNKKREDDVAIEDVLEQFLKDTKPELEIDVKYQOEEDVNVKRRPRMDIETN 600
Db 541 aaeklrsknkremddvaiedevleqflkdtkpeleidvkvqkeedvnrkrprmdietn 600
QY 601 DTSDBAVPSSKISQENEIGKKRELUKEDSLWSAKEISNNNDKLODDSEMLPKKLLITEFR 660
Db 601 dtfsdvaupsskissqeneigkkreikedslwsakeisnnndklqddsemipkklillitefr 660
QY 661 SLVIKNTSRNPSSIINDYCOLKNFKFKKVTYPGAGKLPHIIGGSLIAHHARKNTELE 720
Db 661 slviknstrnpsginddyqglnfkfkfkvtypgagkphiiggsdliiahharknatele 720
QY 721 EWLRQMEVQNHAKESLADDLFRXNPYLKRRR 754
Db 721 ewlrqemevnqhakeesladdlfrxnpylkrrr 754
RESULT 3
AAB92565
ID AAB92565 standard; Protein; 399 AA.
XX
AC AAB92565;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:10768.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX

DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 10768; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 399 AA;
Query Match 52.5%; Score 2046; DB 22; Length 399;
Best Local Similarity 98.7%; Pred. No. 1.7e-148;
Matches 393; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 MWKLLPAAGPAGGEPYRLLTGVEYVVGKNCALIIENDQSLSRNHAVLTANFSVTNLSOT 60
Db 1 mwkllpaagpaggepyrlltgvyyvgrkncalliendqslsrnhavltanfsvtnlsqt 60
QY 61 DEIPVLTLDKNSKYGTFTVNEEKQNGFSRTLKSGDGTGFGVFGSKFRIEYPLVACSSCL 120
Db 61 deipvltldknskylgtftvneekmgngfsrtlksgdgtgfgvfgskfrieyplvacsscl 120
QY 121 DVSGKTALNOAILQLGGFTVNNWTEECTHLVMVSVKVTIKTICALICGRPIVKPEYTFEF 180
Db 121 dvsgktalnqailqlggftvnnwteecthlvmvsvkvtkikticalicgrpivkpeytfef 180
QY 181 LKAVQSKKOPQPTESFYPLDPEPSIGSKNVDSLGRQERKQIFKGKTFIFLNAKOHKKLSS 240
Db 181 lkavqskkppqiesfyppldepsigsknvdlsgqrqerqifkgkctfiflnakghkklss 240
QY 241 AVVFGGGEARLITEENEENHFFLAGPTCVVDGTGINSQTLIPDCOKKWTQSIQMDMLQRQ 300
Db 241 avvfgggearliteeneenehnfflapgtcvvdgtginsqtlipdcqckkwigsimdmqlrq 300
QY 301 GLRPIPEAEIGLAVIFMTTKNYCDPGHPSTGLTKTTPGPSLSQGVSVDEKLMPSAPVNT 360
Db 301 glrpipeaeiglavifmttknycdpgghpstgltkttppgslsqgvsvdeklmpsapvnt 360
QY 361 TTYVADTESQADTWDLSERPKEIKVSKMEQKFRMLSQ 398
Db 361 ttyvadteseqadtwdlserpkeikvskmeqkfrmlsq 398
RESULT 4
AAB62661
ID ABB62661 standard; Protein; 811 AA.
XX
AC ABB62661;

XX
PS Example 4; SEQ ID NO: 32635; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 38 AA;

Query Match 5.3%; Score 205; DB 22; Length 38;
Best Local Similarity 100.0%; Pred. No. 8.5e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 691 VTPGAGKLPHTIGGSDLIHARKNTELEWLROEME 728
Db 1 vtypgagkphilggsdliaharknteleeirlqeme 38
|||||

RESULT 6
AAM73186
ID AAM73186 standard; Protein; 38 AA.
XX
AC AAM73186;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 33492.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR -30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 33492; 658pp + Sequence Listing; English.
XX

CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
SQ Sequence 38 AA;

Query Match 5.3%; Score 205; DB 22; Length 38;
Best Local Similarity 100.0%; Pred. No. 8.5e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 691 VTPGAGKLPHTIGGSDLIHARKNTELEWLROEME 728
Db 1 vtypgagkphilggsdliaharknteleeirlqeme 38
|||||

RESULT 7
AAM33396
ID AAM33396 standard; Protein; 38 AA.
XX
AC AAM33396;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #7433 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488997/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 27; SEQ ID No 33665; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs;
CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
SQ Sequence 38 AA;

Query Match 5.3%; Score 205; DB 22; Length 38;
Best Local Similarity 100.0%; Pred. No. 8.5e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 691 VTPGAGKLPHTIGGSDLIHARKNTELEWLROEME 728
Db 1 vtypgagkphilggsdliaharknteleeirlqeme 38
|||||

RESULT 8
AAM78520
ID AAM78520 standard; Protein; 990 AA.
XX
AC AAM78520;

CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
 CC sequences (ABU01840-ABU16175) and the encoded proteins
 CC (ABBS7737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 1812 AA;

Query Match 4.4%; Score 171; DB 22; Length 1812;
 Best Local Similarity 18.68; Pred. No. 0.001;
 Matches 122; Conservative 115; Mismatches 238; Indels 182; Gaps 26;

QY 200 LDEPSIGSKNDLSGROERKQIFRGKTFIFLNAQKHKLSSAVFGGGEARLITENEEE 259
 Db 5 laetaaaeneeqaddsmeklvfd-----eedgeqmvsgedllgkkslimsdseae 55
 QY 260 HNFFLAGTCVVDGTGINSQTLIPDCKKWTQSDMLQROGLRPIPAEIGLAVIFMTT 319
 Db 56 dg-----sqekplspkdr-----rmmdsqeedam----- 82
 QY 320 KNYCDPOGHPSTGLKTTTPGSLSGQSVSDKLPSPAVNTTIVADTES---EQADTWD 376
 Db 83 -----dggm-----gdsiknsedveee-mvsrpkkvsalidsnddegakekt 126
 QY 377 LSERPKE-IKVSMEQKFRMLSDQAPTVKESCKTSSNNNSMVNTLAKWRIPNYQLSPTK 435
 Db 127 qkeegkpvkskkrsvnqgedqekpvksntkavknktqankaeddqdnpskedkpk 186
 QY 436 LPSITNKSDRASQQQNTNSIRNYFQ----PS-----TKRERDENQE----- 474
 Db 187 vvk-nkkqtnkdgkpqtneeydhghqhekpaktgskklakkqgedkedngteqekk 245
 QY 475 ---WSSCKSARIETSCSLLEQTPA-----TPSLWKNK----- 504
 Db 246 kpspkakksdkakidsimneedagedllmygedqepqkgtkksakktndkseeg 305
 QY 505 --EOHLSENEPEVDNTSNONNLFTDLDLSIVKNSASKSHAAEKLRSNKKRE--MDDVATED 560
 Db 306 edgehgeqkplkikidnidtkedqevipkpmakknkqvlsdeseenqnekvdq 365
 QY 561 EVLQQLFKDKPELEIDVKVQKQEDVNVRK--RPRMDIETNDTFSDEAVP--ESSKI-- 614
 Db 366 hdlkmsesenelmgskdgedqemgeppkakkknqrmde-----sedeipktesekits 420
 QY 615 SQENFICKRELKEDLSWAKEIS--NNDKLQDDSEMLPK----- 652
 Db 421 spknklgvlvdsesepeeteaevspvknklglvdsesepelndpnesageqepmesal 480
 QY 653 -----KLILTFRSRLIVKNTSRNPSGINDDYGO-----LKNFKFKFKVT 692
 Db 481 sreppkavvresakalegmaiqeqqrlhreachinvpyhpkprtllkeflsrrtin 540
 QY 693 YPGAGKPLHIITGGSDLIAHARKNTELEWLRQEME----YONQHAKEESLADDLFR 745
 Db 541 aplatal---aggspmpsrqprksvgl-rmtreeleayakimedrake---ateffk 590

RESULT 10
 ABB67502
 ID ABB67502 standard; Protein: 3257 AA.
 XX
 AC ABB67502;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 29298.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical.
 OS Drosophila melanogaster.

```
QY 391 --QKRLMSODAPTVKESCKTSSNNNGSVNLTAKMRIPNYQLSPTKLPISINKSKDRAS- 447
: | | : : : : : | | | : : | | | : : | | : : |
Db 751 vieivkeleqe--tisevvpqndqssvedqtladkenpvkpsvpkapssskdeppae 808
: | | : : : : : | | | : : | | | : : | | : : |
QY 448 -----QQQTNSIRNYFQSTKKRDEENQEMSSCKSARITSCSLLEQTOPAT 497
: | | : : : : : | | | : : | | | : : | | : : |
Db 809 nlpadqdpieqkt-----pvaknqghdkehneapkaeslsvsdips-----ssvt 855
: | | : : : : : | | | : : | | | : : | | : : |
QY 498 PSLWKNKEHLS-ENEPVDNDSNNLTDTDLKLSIVKNSAKSHAAEKLRSNKKREDDV 556
: | | : : : : : | | | : : | | | : : | | : : |
Db 856 ps----kkrnhsspantpbkkskealealqssvprairlsdkatpqnlres-rskrtlktelt 911
: | | : : : : : | | | : : | | | : : | | : : |
QY 557 AIEDEL-----PQLFKDTRPELEIDVKVQKQED-VNVRKRPRMDI----- 597
: | | : : : : : | | | : : | | | : : | | : : |
Db 912 llmddtmrrsprlgrspaeshsherspmekvtvsklakdlitidkeielkslpda 971
: | | : : : : : | | | : : | | | : : | | : : |
QY 598 -ETND-----PESKISQENIEGKKRELKEDSLWSAKETSNNDKLQ 644
: | | : : : : : | | | : : | | | : : | | : : |
Db 972 setkdvkitkttasdtiitdenpsssktemkklkgkplakkmrtsetevkk--aia 1029
: | | : : : : : | | | : : | | | : : | | : : |
QY 645 DDEMLP-----KLLLETFRSLVKNSTSRNPSGINDXYGOL 682
: | | : : : : : | | | : : | | | : : | | : : |
Db 1030 dsnedipsifskveehltseseqdekeellcpqldctndtleqstaletdeqv 1089
: | | : : : : : | | | : : | | | : : | | : : |
QY 683 KNFX-----RFFKVVYTPGAGKPLHIIGSDLIAHH--ARK--NTELEEWLRQE 726
: | | : : : : : | | | : : | | | : : | | : : |
Db 1090 eekrsnrksrfrirnefktt-----dtlsdhldakkaenasleismrp 1135
: | | : : : : : | | | : : | | | : : | | : : |
QY 727 MEVQNHQA 734
: | | : : : : : | | | : : | | | : : | | : : |
Db 1136 ctletqgs 1143
: | | : : : : : | | | : : | | | : : | | : : |

RESULT 11
ABG27218
ID ABG27218 standard; Protein; 2918 AA.
XX
AC AC
DT DT
TX TX
DE DE
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN W0200175067-A2.
XX
PD 11-OCT-2001.
XX
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS91405.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX PS
XX PS
XX Claim 20; SEQ ID No 57577; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
```

```
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
```

XX Sequence 2918 AA;

Query Match 4.3%; Score 169.5; DB 22; Length 2918;
Best Local Similarity 19.9%; Pred. No. 0.0027;
Matches 185; Conservative 138; Mismatches 293; Indels 313; Gaps 48;

```
QY 2 WKLLPAGPAGGPEYRLLTGVEYVVGRKNCAILIENDOSIRNHAULTAN--FSVTNLQ 59
Db 714 wmlm-----elqkrrglehcegrmtlknpladgg-sshhitvklnelkvsqndq 764
: | | : : : : : | | | : : | | | : : | | : : |
QY 60 TDEIPVLTLDKNSKYGTGVNEEK---MQN---GFSRTLKSGDITFGVSGKFRIEPL 113
: | | : : : : : | | | : : | | | : : | | : : |
Db 765 aiaevlnqlkd--mlanfrgsekycylqnevfglfqklnengvtdgylhslctvr--- 818
: | | : : : : : | | | : : | | | : : | | : : |
QY 114 VACSSCLDVSGKTAALNOAILQLGFTVNNWTEETHLVMWSVKVT-IKTIC-----AL 165
: | | : : : : : | | | : : | | | : : | | : : |
Db 819 -----allqailq-----tedm--lkvyeariteetvcldldkvey 854
: | | : : : : : | | | : : | | | : : | | : : |
QY 166 ICGRPVIVKPE-----YFTEFLKAVQSKQPPQIESFYPPPLDEPSIGSKNVDLSGR 215
: | | : : : : : | | | : : | | | : : | | : : |
Db 855 rcgikkikndlnlkkllatmtkeltelkaqihstqsgqply-dldlgkfgekvttldtr 913
: | | : : : : : | | | : : | | | : : | | : : |
QY 216 QER-----KQIFKGTKF-----IFLNAKHKKLLSSAVVFGGEA--RL 251
: | | : : : : : | | | : : | | | : : | | : : |
Db 914 wqridkqldfrlwdlekqikqlrnyrdnyqafckwlydakrrqdslesmkfgdsntvmrf 973
: | | : : : : : | | | : : | | | : : | | : : |
QY 252 ITEENEEHNFFLAPGTCVVDTGITNSQTLIPDCQKWKIQSIMDMLQROGLRPIPEAFIG 311
: | | : : : : : | | | : : | | | : : | | : : |
Db 974 lneqkn-----lhseisgkrd-----kseevqkiaeicans----ikdyelq 1011
: | | : : : : : | | | : : | | | : : | | : : |
QY 312 LAVIFMTTKNYCDPQGHPTSTGLKTTTPGP-----SLSQGVSVDEKLMPSAPVNTTYVA 365
: | | : : : : : | | | : : | | | : : | | : : |
Db 1012 la-----sy-----tsgetilnlipkrtmigspgvllqe----aadvh-aryi- 1051
: | | : : : : : | | | : : | | | : : | | : : |
QY 366 DTESEQADTWD-LSERPKEIKVSKMEQ-KFRMLSQDAPTWKESCKTSSNNNSMVSNTIAK 423
: | | : : : : : | | | : : | | | : : | | : : |
Db 1052 elltrsgdyvrfisemlksiedklkntkievleeeirlardansencnknkfldnqlk 1111
: | | : : : : : | | | : : | | | : : | | : : |
QY 424 MRIPNYQLSPTKLPISINKSKDRA-----SQQQ 450
: | | : : : : : | | | : : | | | : : | | : : |
Db 1112 yqaecsqfk-aklasleelkrqaeldgksaknldkcygqikeinekitrityeidekr 1170
: | | : : : : : | | | : : | | | : : | | : : |
QY 451 QTNSIRNYF-----OPST 463
: | | : : : : : | | | : : | | | : : | | : : |
Db 1171 rrsvedrfdqkndydqlqargcekenlgwqklesekaikekeyeierlrvllqeggt 1230
: | | : : : : : | | | : : | | | : : | | : : |
QY 464 KKRERDEE-----NOEMSSCKSARIETSCSLLEQTQPATPSLWKNKEOHLSENPEVD 515
: | | : : : : : | | | : : | | | : : | | : : |
Db 1231 rkreyenelakvrnhyneemsnlrn-kyeteinitkttkiei-smqkedsdknlnrq-ld 1287
: | | : : : : : | | | : : | | | : : | | : : |
QY 516 TNSDNNLFTDTDLK-SIVKNSAKSHAAEKLRSNKKREMDVAIETEVLQDFDKTRPEL 574
: | | : : : : : | | | : : | | | : : | | : : |
```

Db 1288 risren-----rdlkdeivrlndsilgateqrrraeenalqgkacgseimqk-----kqhl 1338
QY 575 EIDVK-VQOEEDVNVKRPRMDIETNDT-----FDEAVPESSKISQENE 619
Db 1339 eileikvmqgrsednarhkqsie-eaaktiqdnkneierlkaefgea---krrweyene 1394
QY 620 IGKKRELKEDSLWSAK-----EISNNDKLQDDSEMLPKRLLLT 657
Db 1395 lskvrnnydeefislknqfeteinicktihqitmkeedtsqyraqidnltrenslse 1454
QY 658 EFRSLVIKNS---TSRNPISGINDYQOLKNFKKVKVTPGAGKLPHTIGGSDLIHAHAR 714
Db 1455 eikrl--kntlqttenlrirvesdiq-----qkat-----gsev---sq 1489
QY 715 KNTLEELWLRQ--EMEVOHQHAKESLAD 741
Db 1490 rkqglevelrvqtqmrteesvrykgsldd 1518

RESULT 12
AAW40539
ID AAW40539 standard; Protein; 1128 AA.
XX
AC AAW40539;
XX
DT 28-AUG-1998 (first entry)
XX
DE Mutant C-beta protein nv34qp.
XX
KW C-beta protein; beta antigen; vaccine; group B Streptococcus.

OS Streptococcus sp. strain A909.
XX Synthetic.
XX WO9809648-A1.
XX
PD 12-MAR-1998.
XX
PF 05-SEP-1997; 97WO-US15319.
XX
PR 06-SEP-1996; 96US-0024707.
XX
PA (NAVA-) NORTH AMERICAN VACCINE INC.

XX Blake MS, Tai JY;
XX
DR WPI; 1998-193324/17.
XX N-PSDB; AAV11346.
XX
PT Group B streptococcal C-beta proteins - having amino acid
PT substitutions to reduce immunoglobulin A binding while retaining
PT antigenicity, for use in vaccines

XX Disclosure; Fig 7A-C; 59pp; English.
XX
CC Group B Streptococcus C-beta mutant protein nv34qp has only 10% of
CC the IgA binding activity of the native C-beta protein (see AAW40537).
CC It is encoded by mutated C-beta protein DNA (see AAV11346) in which
CC the Gln-167 and Gln-174 codons of the mature C-beta protein coding
CC region are replaced by Pro codons. These residues are in the
CC IgA binding domain of the C-beta protein. The IgA binding ability
CC of C-beta appears to require dimerisation of the protein. The
CC invention relates to mutant C-beta proteins that have a reduced or
CC eliminated ability to bind human IgA but which retain the
CC antigenicity of the wild-type protein. Such proteins can be used
CC in vaccines for use against group B streptococci. Claimed mutant
CC proteins have mutations in the IgA binding region of C-beta.
XX
SQ Sequence 1128 AA;

Query Match 4.3%; Score 169; DB 19; Length 1128;

Best Local Similarity 19.9%; Pred. No. 0.00072;
Matches 142; Conservative 125; Mismatches 238; Indels 208; Gaps 38;
QY 192 QIESFYPPLDPEPSIGSKNVYDLGR-----QERKQIFKCTFFFLNAKHKKLSSA 241
Db 95 qkneftktdetndsdallelenqfnetnrlhikqneevkdk-----kakqgtlikqs 149
QY 242 VVFGGEARLITEENEHEHFFLAPGTCVVVD--TGITNSOTLIIPDCOKKWTQSIQIMDLQ 300
Db 150 -----dtkvdisnckelnhqspvekmaepkgitned---kdsmlkkiedlrkqaqa 200
QY 301 GLRPIPEAEI--GLAVIFWTTKNYCDPQGHSTGLTKTPPGPSLSQGVSDKLMFSA 358
Db 201 dkkedaevkvreelglfstkagldqeihevkkets-----seentqkvdeh-----a 251
QY 359 NTTTTYVADTESEADTWDLSEPKETKVKME--QK-----FRLMSQDAP 401
Db 252 nslqlaqksleeldkattneqatqvkqnfleaqdklkmqpliketnkvlykamsesle 311
QY 402 TVKESCKTSSNNN--SMVSNLTAKMRIPNYQLSPTK-LPSINKSKDRASO--QOOTNSIR 456
Db 312 qvekelkhseanledlvakskeivreyegklnqsknlpelqleeeahsklkqvvedfr 371
QY 457 NYF----QPSTKKR-ERD-----EENQEMSSCKSARIETSCSLLEQTOPTATPSLW 507
Db 372 kkftseqvtppkkrvkrdlaanenngq-----kieelts-----penitvyegedvk 418
QY 508 LSENEPVDNNSDNNLFTD--TDLKSIYKNSASKSHAAEKLRSNKKREDDVAIEDVLEQL 566
Db 419 ft-----vtaksdsktldfisdL--ltkynpsvs---drisnkyktntdnkhaieftiknl 469
QY 567 -----FRDTRK---PELEIDVKVQOEEDVNVKRPRMDIETNDTFSDEAVPESS 612
Db 470 klnesqvtlikakddsgnvvektftitvqkkeek-qvpktp-----eqkdskteekvpqep 524
QY 613 KISQENEI-----GKKRELK-----EDSLW-SAKE-----IS 638
Db 525 ksndknqlgeliksaggeleklekaikelmegpeipsnpeyigkswesqkepigeait 584
QY 639 NNDKLQDDSS-----EMLPKKLLL-----TEFRSLVI 664
Db 585 sfkkliigdsesskytyehyfnkyksdfmnyqlhaqmemltrkvvqymnkypdnaelkk-if 643
QY 665 KNSTSRNPISGINDYQOLKN-----FKKF-----KKVTYPGAGKLP 702
Db 644 esdmkrtk---ednygslendalkgyfekyfltpfnkikivddldkkveqdpapape- 699
QY 703 IGGSDLIHAHARKNTELEEWLRQEMEVQNHQ--AKEESLADDLFRYNPYLKRR 753
Db 700 --nsemdqakekakiaavskymskvldgvhqlqkknhsktivdlfkeleakqg 750

RESULT 13
AAW84461
ID AAW84461 standard; Protein; 1128 AA.
XX
AC AAY84461;
XX
DT 25-JUL-2000 (first entry)
XX
DE Amino acid sequence of a C-beta protein.
XX
KW C-beta protein; bactericidal; gram positive bacteria; vaccine;
KW immune response.
XX
OS Streptococcus agalactiae.
XX
PN WO200015760-A1.
XX
PD 23-MAR-2000.
XX
PF 17-SEP-1999; 99WO-US21643.
XX


```
XX SQ Sequence 808 AA;
Query Match
Best Local Similarity 4.3%; Score 168; DB 22; Length 808;
Matches 125; Conservative 112; Mismatches 204; Indels 148; Gaps 32;
QY 252 ITEEN-----EEHN-----FFLAPGTCVVDTCI-TNSOTLIPDCQKWKIQSIMDMLQK 300
Db 109 lteenfrlqaehdqakelfllrktleemrlretqkql--nadesikkllemlqsk 166
QY 301 GL-----RPIPAE--IGLAVIFMTKN-----YCDPQGHPSGLTKTTPGPSL 342
Db 167 gipskslednertirmaeaeqsvshvildqkekenihlreelhrsqglq---pepak 223
QY 343 SOGVSVDEKLMPAPVNTTYYVADTSE-----QADTWDLSEPRKEIKVSKMEQKFR 394
Db 224 tkalgtviemkdkiaslernirdlelqmkangvlnedreeikqlevykhskfm 283
QY 395 -----MLSQDAPTAKES-----CKTSSNNMSVSNPLAKMRIPNYQLSPTKLPISINK 441
Db 284 ktidqkqel-sktesellalqtkletlsnqpsackqhiwlvk-----eslta 331
QY 442 SKDRAS-QQQQNSIR-----NYFPSTKK-RERDEEN-----QEMSSCKSAR 482
Db 332 keqraailqtevdaalrleekesfinkktkqlqlditeekgtlageirdmkmlevkerk 391
QY 483 IETCSLLEQTQAPPSLWKNEQHLs-----ENEPVDTNSDNNLFTDTDLKSIVKNSASK 538
Db 392 lnvlqklienlqeql-----rkdkgkltlnldrvkslqtdssn---tdtalatleeealsek 444
QY 539 SHAAEKLRSNKREMDVAIEDVLEOLFQKDKPELEIDVYKQKOEEDVNVKRPRMDI- 597
Db 445 erlierlk--eqrerdd-----rerleesfrkenkdleknalqaelitekesslidlk 498
QY 598 ETNDTFSDEAVPESSKI-SQENEIGKKRELKEDLSWASKEISNNDKLODDSEMLPKLLL 656
Db 499 ehassiasaglkrdsklslelaieqke--ecsklea-qlkkahnieddsrmpn----- 550
QY 657 TEFRLSVIKNSTRNPSGINDYD-----OLKNFK--KFKKVTPYGAGKLP 700
Db 551 -efadqil--kqidkeasyrdecgkaqaevdrllellkevenekndkdkkiaelesltir 607
QY 701 HIIGSDLIAH-----HARNTLEEMLRQEMEYONHAKESLADD 742
Db 608 hmkdgnkvvanlkhnqglekknaglleevrr-----redsmadn 647
RESULT 15
AAB93691
ID AAB93691 standard; Protein; 832 AA.
XX
AC AAB93691;
XX
XX 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:13253.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
```

```
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
DR
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 13253; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 832 AA;
Query Match 4.3%; Score 168; DB 22; Length 832;
Best Local Similarity 21.1%; Pred. No. 0.00055;
Matches 133; Conservative 100; Mismatches 250; Indels 148; Gaps 27;
QY 185 QSKKQPQIESFY---PPLDEPSIGSKNVLDLSCRQERKQIFKGTIFLNAKQHKLLSA 241
Db 142 eakktgptifspfcsmppi-fptvgkdkvn-----nlladpenivtyknrernamdfs 193
QY 242 VVFGGEARLITEENEENHFFLAPGTCVVDTCITNSOTLIPDCQKWKIQSIMDLQ--- 298
Db 194 svf--spsfqipevleceenyfcsf---vnsngklsdlsiltnepvk-----rdsldashy 242
QY 299 ---RQGLRPIPEAEIGLAVIFMTKNKYCD-----PQGHPSGLTKTTPGPSLSQ 345
Db 243 sgvsqnir--skaqi-lallkssesssceeInsemtehfpkqpggsklatkpkylqq 299
QY 346 VSDEK-----LMPAPVNTTYYA-DTSE-QADTWDL 377
Db 300 eecaemkstenlyyqhgsentmrnkswamyfssqssphsthsstvdgndtcrkpaqddv 359
QY 378 SERPKKIKVSKMEQKFRMLSQDAPTVKESCKTSSNNMSVSNPLAKMRIPNYQLSPTKLP 437
Db 360 nsnlklsiqkiiqfvetyaeerk--kynvdqsvgndpswqevkleipsf----- 409
QY 438 SINKSDRASQQQQTNSIRNYFQPTKKRERDEE---NQEMSSC-----KSARIETS 486
Db 410 ----nessslqvctssaendgilssedlqednkkipfnqndkqcgkivlikenaqevnt 464
QY 487 CSLLLEQTQAPPSLWKNEQHLSENEPVDNTNSDNNLFTDTDLKSIVKNSASKSHAAEKL 546
Db 465 cgtleke-----yeqssslpelkhlqiessnmrsrisdditdmises----- 506
```

[illegible]

Search completed: August 15, 2002, 01:40:35
Job time: 6225 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 00:14:45 : Search time 42.3 seconds
(without alignments)
1712.798 Million cell updates/sec

Title: US-09-837-602-2
Perfect score: 3899
Sequence: 1 MWKLLPAAGPAGGEPYRLLT.....KEESLADDFRYNPKRRR 754

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3896	99.9	754	2 T00393	Nijmegen breakage
2	227.5	5.8	1460	2 T00095	hypothetical prote
3	215.5	5.5	1526	2 A45605	mature-parasite-in
4	185.5	4.8	3147	2 T18674	hypothetical prote
5	185.5	4.8	3488	2 T34418	hypothetical prote
6	179.5	4.6	1790	2 S67593	transport protein
7	177	4.5	1211	2 T27522	hypothetical prote
8	174	4.5	911	2 S51441	hypothetical prote
9	173	4.4	1938	2 I49464	alpha cardiac myos
10	172.5	4.4	1038	2 JC5497	claustrin - chick
11	171	4.4	465	2 A02986	myosin alpha heavy
12	171	4.4	1939	2 I49175	myosin heavy chain
13	170	4.4	2464	1 QRMSP1	microtubule-assoc
14	169.5	4.3	1478	2 S20117	protein kinase BCK
15	169.5	4.3	2287	2 T21312	hypothetical prote
16	169	4.3	1229	2 T25697	hypothetical prote
17	168.5	4.3	1125	2 E90598	membrane nucleas
18	168.5	4.3	1233	2 S56271	hypothetical prote
19	168.5	4.3	1939	1 A46762	myosin alpha heavy
20	168.5	4.3	3672	2 T23433	hypothetical prote
21	168.5	4.3	3704	2 T37316	probable laminin a
22	167.5	4.3	853	2 T51505	hypothetical prote
23	167.5	4.3	2094	2 S33124	tptr protein - huma
24	166	4.3	1359	2 T34036	hypothetical prote
25	166	4.3	1935	1 A37102	myosin beta heavy
26	165.5	4.2	1272	2 C90593	hypothetical prote
27	165	4.2	667	2 A40713	cyclicin I - bovine
28	165	4.2	790	2 S67803	probable membrane
29	165	4.2	1410	1 A57013	early endosome ant

30	165	4.2	1938	1 S06005	myosin alpha heavy
31	164.5	4.2	1392	2 A43336	microtubule-vesicl
32	164.5	4.2	2677	2 A38194	desmoplakin I - hu
33	164	4.2	915	2 S54485	CES1 protein - yea
34	164	4.2	1381	1 S45781	probable calcium-b
35	164	4.2	1738	2 T18837	interaptin - slime
36	163	4.2	952	2 T18837	hypothetical prote
37	162.5	4.2	539	2 S40900	FUN17 protein - ye
38	162.5	4.2	1427	2 S22695	restin - human
39	162.5	4.2	1957	2 T38077	hypothetical colle
40	162	4.2	1263	2 T13465	hypothetical prote
41	162	4.2	1312	1 BWBYDL	RAD50 protein - myc
42	162	4.2	1365	2 T30822	limp1 protein - Myc
43	162	4.2	2748	2 S57976	nuclear migration
44	161.5	4.1	1087	2 T30330	gelsolin-related p
45	161	4.1	1928	2 S46773	myosin heavy chain

ALIGNMENTS

RESULT 1
T00393

Nijmegen breakage syndrome protein NBS1 - human
N:Alternate names: cell cycle regulatory protein p95
C:Species: Homo sapiens (man)

C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 20-Jun-2000
C:Accession: T00393
R:Matsuura, S.; Tauchi, H.; Nakamura, A.; Kondo, N.; Sakamoto, S.; Endo, S.; Smeets,

Mature Genet. 19, 179-181, 1998
A:Title: Positional cloning of the gene for Nijmegen breakage syndrome.
A:Reference number: Z14144; MUID:98282099

A:Accession: T00393
A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA
A:Residues: 1-754 <MAT>

A:Cross-references: EMBL:AB013139; FIDN:BAA28616.1
C:Genetics:

A:Gene: NBS1
A:Map position: 8q21

A:Introns: 13/1; 57/3; 107/2; 160/3; 195/2; 234/3; 299/2; 332/1; 375/2; 466/2; 615/3,
C:Superfamily: human Nijmegen breakage syndrome protein NBS1

C:Keywords: DNA repair

Query Match 99.9%; Score 3896; DB 2; Length 754;
Best Local Similarity 99.9%; Pred. No. 7.4e-201;
Matches 753; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWKLLPAAGPAGGEPYRLLTGVEYVVGKRNCAILINDQSIISRNHVLTFANFVTLNSQT 60
|||||

Db 1 MWKLLPAAGPAGGEPYRLLTGVEYVVGKRNCAILINDQSIISRNHVLTFANFVTLNSQT 60
|||||

QY 61 DEIPVLTLDKNSKYGFVNEEKQNGFSRTLKSGDGIITGVFGSKFRIEYPLVACSSCL 120
|||||

Db 61 DEIPVLTLDKNSKYGFVNEEKQNGFSRTLKSGDGIITGVFGSKFRIEYPLVACSSCL 120
|||||

QY 121 DVSGKTKALNQAILQLGGFTVNNWTECTHLVWVSVKVTIKTICALICGRIVKPEYFTEF 180
|||||

Db 121 DVSGKTKALNQAILQLGGFTVNNWTECTHLVWVSVKVTIKTICALICGRIVKPEYFTEF 180
|||||

QY 181 LKAVQSKQPPQIESFYPPPLDEPESIGSKNVLDLSGRQERKQIFKGTFFIFLNAKQHKLLSS 240
|||||

Db 181 LKAVESKKQPPQIESFYPPPLDEPESIGSKNVLDLSGRQERKQIFKGTFFIFLNAKQHKLLSS 240
|||||

QY 241 AVVFGGGEARLITEENEEHNFFLAPGTCVVDGTGITSQTLIPDCQKWTQSIIMDLQRO 300
|||||

Db 241 AVVFGGGEARLITEENEEHNFFLAPGTCVVDGTGITSQTLIPDCQKWTQSIIMDLQRO 300
|||||

QY 301 GLRPIPEAEIGLAVIFMTTKNYCDPOGHPSTGLKTTTPGPSLSQGVSDSKLMPSPVNT 360
|||||

Db 301 GLRPIPEAEIGLAVIFMTTKNYCDPOGHPSTGLKTTTPGPSLSQGVSDSKLMPSPVNT 360
|||||

Jun 19

C:Keywords: surface antigen; tandem repeat

Query Match 5.5%; Score 215.5; DB 2; Length 1526;
Best Local Similarity 20.5%; Pred. No. 0.0012;
Matches 131; Conservative 118; Mismatches 273; Indels 117; Gaps 27;
QY 168 GRPIVK-PEYFTFLKAVQSKKOPQIESFYPLDPSIGSKNVDLSGRQERKQIFKGKT 226
Db 633 GNDKVGKPEIITEVKE-EIKK---QVED-----GIKENDTEGNDK-----VKGPE 674
QY 227 FIFLNAQH--KKLSAVVFGGEA-----RLITEENEEENFFLAPGTCVVDGIGTN 277
Db 675 IITEEVKEEIKQVEGIEKENDTEGNDKVGKPEIITEEVKEEIKQVEGVKENDTESKD 734
QY 278 ---SQTLPDCKKWTQSIQMDLQROGLRPIPEAEIGLAVIFMTTKNYCDPOGHPSTGLK 334
Db 735 KVIGQIITEEVKKEIEKQEKGNKENILEIKDIVIQEVEIIEEVKKVKKK--VEKGIK 792
QY 335 ---TTTPGPSLSQGVSD-E-KLMPSPAVNTTYYADTES-----EQADTWDLSEPK-E- 383
Db 793 ENHTESKDQVIGQIITEEVKEEIEKQVEGIEKENDTESKDKVIGEEVKGVDVNEEGPEN 852
QY 384 -IKVSKMEQKFRMLSQADATVYKESCTSSNNNSWNTLAKMRIPNYQLSPTKLPSINKS 442
Db 853 KDKVTQKQEVKKEVKKVKKVKKRKNRKNKNDKNDVIGKEIMKE---DVNEKDTANKD 909
QY 443 KDRASQOQOQNSTRNYPQFSTKRRDEENOEMSSCKSARIETSCSLLEQ-----TOPAT 497
Db 910 KE-IEQEKKEEYKE--KEEYKEKEEYKEEYKEEYKEEYKEEYKEEYKEEYKEDTESKD 966
QY 498 PSLWKKE-OHLSNEFPVDTNSNNLFTDPLKLSIVKNSASKSHAAEKLRSNKKREMDVY 556
Db 967 KETEKEKEEYKEEYKEDTKNDKQVIGQIITEEIKKEVKK---RVKRRNNKNENKDNV 1023
QY 557 AIDEVLEQL-FKDTPELEIDVKKQOEEDVNVRRPRMDIETNDTFSDEAVPESKIS 615
Db 1024 IVOEIMNEDVNEKDTANKDKV-IEQEKKEEVK-----BKEEYKEEYKEEYKEEVK 1073
QY 616 QNEIGKRELKRDLSWS-----AKETSNNKLDQDSE-----MLPKKLLLTFRSLVKN 666
Db 1074 EKEEYKEEYKEDTESRDNVIVQELMNEDEVNEKDTESKDKMIGEVIEEYKEEVKKR 1133
QY 667 STSRNPSGIN-----DDYGQLKNPKKF-KKVTYPGAGKLPHIIG 704
Db 1134 VNKEVKNVRNRNKRKNDKVDIEQIEVSEVNEKDKTKNDKKTGKRVKKP----- 1183
QY 705 GSDLIAHARKNTLEEWLRQEMEVQNHAKESLADDL 743
Db 1184 -----IDCKKEREVOEIMNEDVNEKDTESKDKMIGEV 1217
RESULT 4
T18674
hypothetical protein T04F3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T18674; T24464
R:White, S.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19004
A:Accession: T18674
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3147 <W12>
A:Cross-References: EMBL:274026; PIDN:CAA98419.1; GSPDB:GN00023; CESP:T04F3.1
A:Experimental source: clone B0240
R:Kershaw, J.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z19894
A:Accession: T24464
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-3147 <W12>
A:Cross-References: EMBL:272513; PIDN:CAA96672.1; GSPDB:GN00023; CESP:T04F3.1
A:Experimental source: clone T04F3
C:Genetics:
A:Gene: CESP:T04F3.1
A:Map position: 5
A:Introns: 338/3; 417/3; 1957/2; 2358/3; 2400/1; 2529/2; 2585/1; 2700/3; 2744/1; 287/
Query Match 4.8%; Score 185.5; DB 2; Length 3147;
Best Local Similarity 19.0%; Pred. No. 0.12;
Matches 156; Conservative 144; Mismatches 295; Indels 225; Gaps 40;
QY 60 TDIPVLTLK-----DNSKYGTGTF---VNEEKMQGFSRTLKSGDGIITFGVGSKFRI 108
Db 651 TENQPETTLELDVEKHEIDSSKVSTSTINLNDSEMETRNTNDSKDSFDDVNPQRNLVV 710
QY 109 EY---EPLVACSSCL-----DVSGKALNQAAIL-----QLGGFTVNNWT 144
Db 711 EIPFSEPRVTSTATIQLESDDVAGENSENKRPRVISMRSKEIAKKEKDAQRSGFVIIPHS 770
QY 145 EECTHLV---MVSVKVTIKTI---CALICGRPIVKPEYFTEFLKAVOSKKOPP-----Q 192
Db 771 KE--HILDESNIEMDDVFNTPHDQCRV-----PDADSSKHTSSRESVSTVTIN 820
QY 193 IESFYPLDPSIGSK-NVDLSGRQER---KQIFK--GKTFIFLNA-----KQ 234
Db 821 LDNVF-PTEPKLVAKDNCIEAEERIGKRIRKIFQERTTGEQEISKNSEPTEDEMPDEKD 879
QY 235 HKLLSSAV-----VFGGGEARLITEENEEHNFPLAGTCVVDGTITNSQTLIPDCQKK 288
Db 880 HR--TSAVSDLDLKVFFQGTAK--KPENDEFDE-----KIKRGIAEFER---SKQEK 924
QY 289 WIO-STIMDLQROGLRPIPEAEIGLAVIFMTTKNYCDPOGHPSTGLTTPPGSPLSQGV 347
Db 925 EVORSVGAETSHSGKHIFDESNIEMDDVFNVSQYKKEDE-----KLSTPERIVEEVS 977
QY 348 VDEKLMPSPAVNTTYYVADTESQADTWDLSEPKETIKVSKMEQKFRMLSQADATVYKESC 407
Db 978 -----TATMNLNIIISAS-----GIATREENTNVLSEERIQ-----KRVEEFK 1016
QY 408 KTSNNNSWNTLAKMRIPNYQLSPTKLPSINKSKDRASQOQOQNSTRNYPQFSTKRE 467
Db 1017 KTTENLEIQEVVLTKEEVDNSDVKEHRTSAVNIIDLEKV-----FIHGSSKKPK 1065
QY 468 RDEE-----NOEMSSCKSARIETSCS---LLEOTQPATPSLWKNKEOHLSENE 512
Db 1066 NDDEKIRGTAEPERTKQEKAOBSTVIETSSQNSRIFESSISMDDVFNNSLHNSQVS 1125
QY 513 PVDTNSDNN--LFTDIDLKSIIVKNSASKSHAAEKLRSNKKREMDVAVI-----EDE 561
Db 1126 EITEASDPDLVLTSTTFHNVIEEKIDDD--VTKTDNSVEEKEEQVRLRIDEFKRPTBEQ 1183
QY 562 VLEQLFKDTKPELEIDVKKVQKQEDNVV-----RKPRMDIETNDTFSDEAVPESS 612
Db 1184 NLOKEFLTKEBEYSYKMEENRTSAVSDLDKVPDQSKKETTVSNTEXDEIK-RGIAEFE 1242
QY 613 KIQEENEIGKK-----RELKEDLSWAKETSNNDKLQDDSEML--PKKLLLTFRS 661
Db 1243 RSKQEKVQRSGVAETSHSGKHIFDESNIEMDDVFNVSQYKKEKLSSEPRTVEPEVST 1302
QY 662 LVINKNSTRNPSGI-----NDYICQ-----LKNFKKPKVVTYPGAGKLPHIIGSDLIAH 711
Db 1303 ATWNLNIIIFASGIATREENTDVLSEERIQKRVEEKTKT----- 1343
QY 712 HARKNTELE-EWLRQEMEVQNHAKESLA-----DDLDF 744
Db 1344 ---ENLEIQEVVLTKEEVDNSDVKEHRTSAVNIIDLDVDF 1380
RESULT 5
T34418
hypothetical protein F12F3.3 - Caenorhabditis elegans

QY 738 ----SLADLER 745
Db 1385 EKINTLEDELIR 1396

RESULT 7
T27522
hypothetical protein ZC373.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T27522

R:Kershaw, J.
submitted to the EMBL Data Library, April 1995
A:Reference number: Z20382
A:Accession: T27522
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1211 <WIL>
A:Cross-references: EMBL:Z49131; PIDN:CAA88976.1; GSPDB:GN00028; CESP:ZC373.4
C:Experimental source: clone ZC373
C:Genetics:
A:Gene: CESP:ZC373.4
A:Map position: X
A:Introns: 20/1; 53/2; 80/3; 122/2; 169/3; 198/3; 236/2; 286/2; 337/3; 381/3; 402/3; 445/3

Query Match 4.5%; Score 177; DB 2; Length 1211;
Best Local Similarity 20.7%; Pred. No. 0.098;
Matches 171; Conservative 137; Mismatches 317; Indels 200; Gaps 42;

QY 3 KILPAGPAGBP-----YRLTGVYVGR-----KNCAILIENDQISRRNHAVITANF 52
Db 455 KLVETDGDAGLRPPPELMTKRSTGSGESLLAAVKVLKLEV-VENGNSPKRK----- 505

QY 53 SVTNLSQTDIPIVLTKDMSKYGTFVNEEKMONGFRTILKSGDIGTFVFGSKFRIEYEP 112
Db 506 SLCSVKEEIP--AKKQTEGYKTL---ETTENG----- 534

QY 113 LVACSSCLDVSGKTA-LNAQILQLGGFTVNNMTEECTHLVMVSVKVTIKTICALICGRPI 171
Db 535 -----DIRKSAPVKQDTL-----TCQDMSKVTPLQLETPAPNGEIRKKSASAKEKS 581

QY 172 VRPEYFTEFLKAVSQKKPPQIESFYPLDEPSIGSKNYDLGSRQKQIFRGKTFIFLN 231
Db 582 VRPEVIR-KKSSASVKETVELSVVPVKLELAS--IDEGEKLAVK----- 627

QY 232 AKQHKLLSNAVFGGEARLITEENEENHFFLAGPTCVDTGITSQTLIPDCQKWTQ 291
Db 628 ----KKKTTVTAEIEKPKSVIESKEE-----LVEKPKLLNGTLVSDALEK-RR 672

QY 292 SI-----MDMLQRLQRLPIPEAEIGLAVIFMTTKNYCDPOGHPSTGLKTTTP-----G 339
Db 673 SLGSKRASQDVGRLPLVDFSPKIVLTDKEPKESKTKPAEPLKEIVPETKKPARQ 732

QY 340 PLSQGVSVDEKLMPSPVNTT---TYVADTSEQADTWDLSEPKIEIKVSKMEQKFRML 396
Db 733 ALRKRDKSLDD-ILNAAPSTTILSKSTVTSVKHSVYET--IHPSGEBIIKSEKVLTKKKS 789

QY 397 SQD-----APTVECKSTSSNNNSVNTLAKMRIP-----NQQLSP 433
Db 790 KDKSDIDRLPVKIIEDKASLKLKLDVVTFFESKPVKTKSVIKVL-IPEDANLIKDEKRP 848

QY 434 TKLPSINKSKRASQOQQTNSIRNYFPQSTKKRER--DEENQEMSSCKSARIETSCSLLE 491
Db 849 TKI-VMTSEKQATTVTQTTTSVKRTTKSEKKVEHTADGKSVESQKNSQKDD----VK 903

QY 492 QTQPATPSLWKNKEQHLSENPVDNTSDNNLFTDPLKSTIVKNSASKSHAAEKLRSNKKR 551
Db 904 VSQVTT-----KKEEDSTQAP-----TLTVKKNVVKQTAEKSTSEKHKHTENK 948

QY 552 EMDVDAIEDEVLEQLFKD--TKPELEIDVKVQKOEEDVNVKRPRMDTETNDTDSDEAVPE 610
Db 949 EOEIRIIKKSSTLKD--KDAQVSLQOSIKVNGEKHKESKVGIGTKTVINKLELKEETIN- 1005

QY 606 EAVPES-----KISQEN-----EI---CKKRELKEDSLWSAKEISNNDKLQDSEMLP 651
Db 1006 -ANPKSSSLKATAKVGTGETTKHLEISTDGLKGLVTKDTSKM--LSPDSSSDTSSKN 1062

QY 652 KLLLFTEFRSLVTKNSTSRNPSSINDDYQOLKNFKPKFKVYTPGAGKLPHIIGSDLIAH 711
Db 1063 VRILLTDEQS--VKSSVRSHPSVSTRDSEER-----KKYRFAGDVGTPTKLNPPPLFGN 1115

QY 712 ---HARKNTELEWLRQEMEVQNH-----AKEESLADDLFRYNP 748
Db 1116 GLOKMRSESSLHEKI-NGLTTKKSHDDVSLARKTSLFTDI-EYQP 1158

RESULT 8
S51441
hypothetical protein YLR309c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein L2142.5
C:Species: Saccharomyces cerevisiae
C:Date: 23-Feb-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C:Accession: S51441
R:Pauley, A.
submitted to the EMBL Data Library, November 1994
A:Description: The sequence of S. cerevisiae cosmid L2142.
A:Reference number: S51437
A:Accession: S51441
A:Molecule type: DNA
A:Residues: 1-911 <PAU>
A:Cross-references: EMBL:U17247; NID:g577216; PIDN:AAB67359.1; PID:g577221; MIPS:YLR309c
C:Genetics:
A:Gene: SGD:IMH1
A:Cross-references: SGD:S0004300; MIPS:YLR309c
A:Map position: 12R

Query Match 4.5%; Score 174; DB 2; Length 911;
Best Local Similarity 21.2%; Pred. No. 0.097;
Matches 138; Conservative 113; Mismatches 275; Indels 126; Gaps 31;

QY 179 EFLKAVQSKKQ-----PPOIESFYPLDEPSIGSKNYDLGSRQER---KQIPKGTTFIF 229
Db 141 EFLKQNYKNDIDDLKKMEALNIEDTVQ-KEKNDTVSGLREKIVALENILEER--- 196

QY 230 LNAQHKLLSNAVFGGEARLITEENEENHFFLAGPTCVDTGITSQTLIPDCQKWK 289
Db 197 -EAKQKEEVS-----ISELKEE-----LAIKHSLSDSRMKITELQN- 234

QY 290 IQSIMDLQRLQGLR-----PIPEAEIGLAVIFMTTKNYCDPOGHPSTG-----LKTTP 338
Db 235 LSSKSTIMEBKSSSELAELNITLKEKPKLSELEKMKELPKAISHQNVGNRRKKNRK 294

QY 339 GPSLSQGVSV--VDEKLMPSPVNTTYVADTSEQADTWDLSEPKIEIKVSKMEQKFRM 395
Db 295 GKXKNGGITGDISEBETVDNSINTBEYDKLENLQ---ELOEKYKDCB--DWQKYED 348

QY 396 LSQDAPTKR-----SCKTSSNNNSVNTLAKMRIPNYQLSPTK---LPSINKSKDR 445
Db 349 IEAELDKAELENSQLEKSAKELETNLTELIDTKSLKKNSELEEVDRMLRTVGNELVD 408

QY 446 ASQOQQTNSIRNYFPQSTKKRDERNQ-----EMSSCKSARIETSCSLLEQTQATPS 499
Db 409 AKDEIKESSKQNEEVKTVKLEDDLRLHKHKNATIEAYEAKNTELRSLKIELLSKVEHLKN 468

QY 500 LW--KKEQHLSENPVDNTSDNNLFT-----DTDLKSTIVKNSASKSHAAEKLRSNKKR 551
Db 469 LCTEKEKEQTQSNQKVAKLENEEISQLYEKSNTITKELTSURT SYKQKEKTVSLVEEQVK 528

QY 552 EMDVDAIEDEVLEQLFKD--TKPELEIDVKVQKOEEDVNVKRPRMDTETNDTDSDEAVPE 610
Db 529 FSEQDKVAEKSTQLRKDHAKISNRDL-LKKENETLH-----NDTAKNSNYEYELKE 581

QY 611 SSKISQENETGKKRELKEDSLWSAKEISNN--DKLQDSEML-----PKLL-----L 656

Db 582 NGKLSERLNIQE---KYNTLQNVKNSNEHIDSIKQCEBLNVKLESTKKILSLSEDEL 638

QY 657 TEFRLSVIKNSTGRN-----PSGINDDYQLKNFKFKKVTY--PGAGKLPHIIGSD 708

Db 639 NEYANIYQDRTREANTLRLRVSDSQTDSSKQKELE--NKLAVLTDEKNKLE-----AEL 691

QY 709 IAHHARKNTELEWLRQMEVQNO-HA---KEESL---ADDLFRYNPYLKR 753

Db 692 DLQTSRKATELQEMKHTVTTELKSEIHALKLEEGKSEVDALHVNNDIKRK 743

RESULT 9

I49464

alpha cardiac myosin heavy chain - mouse

C;Species: Mus musculus (house mouse)

C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Feb-2001

C;Accession: I49464; I49463; I49462; I49604

R;Quinn-Laquer, B.K.; Kennedy, J.E.; Wei, S.J.; Beisel, K.W.

Genomics 13, 176-188, 1992

A;Title: Characterization of the allelic differences in the mouse cardiac alpha-myosin H

A;Reference number: A38207; MUID:92250040

A;Accession: I49464

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1938 <RES>

A;Cross-references: GB:M76601; NID:g191623; PIDN:AAA37162.1; PID:g191624

A;Accession: I49463

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-193, 'D', 195-837, 'S', 839-955, 'N', 957-1938 <RE2>

A;Cross-references: GB:M76600; NID:g191621; PIDN:AAA37161.1; PID:g191622

A;Accession: I49462

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1938 <RE3>

A;Cross-references: GB:M76599; NID:g191619; PIDN:AAA37160.1; PID:g191620

A;Accession: I49461

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-544, 'A', 546-1938 <RE4>

A;Cross-references: GB:M76598; NID:g191617; PIDN:AAA37159.1; PID:g191618

R;Gulick, J.; Subramaniam, A.; Neumann, J.; Robbins, J.

J. Biol. Chem. 266, 9180-9185, 1991

A;Title: Isolation and characterization of the mouse cardiac myosin heavy chain genes.

A;Reference number: I49604; MUID:91225025

A;Accession: I49604

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-67 <RE5>

A;Cross-references: GB:M62404; NID:g192609; PIDN:AAA37424.1; PID:g192610

C;Superfamily: myosin heavy chain; myosin motor domain homology

C;Keywords: ATP; cardiac muscle; heart; nucleotide binding; P-loop

F;88-768/Domain: myosin motor domain homology <MMOT>

F;178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 4.4%; Score 173; DB 2; Length 1938;

Best Local Similarity 22.7%; Pred. No. 0.3;

Matches 92; Conservative 61; Mismatches 122; Indels 130; Gaps 17;

QY 365 ADTESEQADTWLSERPK-----EIKVSKMEQKFRMLSDQAPTVKESCTSSNNNSWS 418

Db 845 AETEKEMANNKEBFRGVKDALEKSEARKKELEKMWSLQEKNDLQVQAEQDN----- 899

QY 419 NTLAKMRIPNYQLSPTKLPINSKDRASQQQNTNSIRNYFQPSKKR---ERDEENQEM 475

Db 900 -----LNDABERCQ-----LLKNKIQLEAKVKEMTERLEDEEM 934

QY 476 S---SCKSARIETSCS--LLEQTOPATPSLWK--NKEQHLSENEPVDNTSDNNLFTDLDKS 530

Db 935 NAEITAKKRKLEDECSLKKDIDDLLETLTAKVEKEKHATENK-----VKN 979

QY 531 IVNSASKSHAAEKLRSNKK-----REMDDVAIEDVEQLFKDTPKPELIDVKYQK 583

Db 980 LTEMAGLDBIIAKLTKEKKALQBAHQOALDDIQAEDEKQVNTL---TKSKVKLEQQVDDL 1036

QY 584 EEDVNVKRRMDTETNDTFESDEAVPESSKISQENIEGKKRELKEDSLSAKAISN--ND 641

Db 1037 EGSLEQEKKYRMDLER-----AKKLEGLDKLTQESIMDLND 1074

QY 642 KLQDSEMLPKKLLTTEFRSLVITKNSTRNPSSGINDYQ-OLKNFKPKKVTYPGAGKLP 700

Db 1075 KLQ-----LEEKLKKKEF-----DISQNSKTEDEQALALQKKLKE----- 1112

QY 701 HIIGGSDLIHAHARKNTELEEWLRQMEVQNOHAKESLADDLFR 745

Db 1113 -----NQARIEEELEEAE---RTARAKVEKLRSDLSR 1143

RESULT 10

JC5497

claustrin - chicken

N;Alternate names: keratan sulfate proteoglycan

C;Species: Gallus gallus (chicken)

C;Date: 07-Jul-1997 #sequence_revision 12-Sep-1997 #text_change 21-Jul-2000

C;Accession: JC5497; PC4334; S37561

R;Burg, M.A.; Cole, G.J.

J. Neurobiol. 25, 1-22, 1994

A;Title: Claustrin, an antiadhesive neural keratan sulfate proteoglycan, is structura

A;Reference number: JC5497; MUID:94157526

A;Accession: JC5497

A;Molecule type: mRNA

A;Residues: 1-1038 <BUR1>

A;Cross-references: EMBL:X67778; NID:g406318; PIDN:CAA47988.1; PID:g406319

A;Accession: PC4334

A;Molecule type: protein

A;Residues: 79-83;299-412;485-502 <BUR2>

A;Experimental source: brain

C;Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the ne

C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate

F;287-270/Region: cell attachment (R-G-D) motif

F;112,213,490/Binding site: carbohydrate (Asn) (covalent) #status predicted

F;152,249,440,793,820/Binding site: chondroitin sulfate (ser) (covalent) #status pred

Query Match 4.4%; Score 172.5; DB 2; Length 1038;

Best Local Similarity 18.6%; Pred. No. 0.14; 326; Indels 309; Gaps 42;

Matches 176; Conservative 135; Mismatches 326;

QY 6 PAAGPAGGEPYRLLTGVEY---VVRKNCAILLIENDQ-----SISRNHAVLTANFSVTN 56

Db 19 PAA-PPSPSLSHRFLLDSKFYLLVWIGE-----LVTEEHLRAIANIERGIRSDMTNLTEN 72

QY 57 LSQTFDEIPVITLKDNSKYGTFTVNEEKWONGFSRTLK-----SGDGIITFGVFGSKPRI- 108

Db 73 LDQ--ELKLFVSRHSARFSPEVRGQKILHHRSDVLETVVLINPDEAVS-----TEVRLM 125

QY 109 -----EYEPLVACSSCLDVSKTALNQAILQLGGFTVNNWTEECTHLVMVSVKVTI---- 159

Db 126 ITDAARHKLLVLTGQCFENTG-----ELILQSGSFQNFIEITFDQIEGELLSTTHPAN 180

QY 160 KTCALALC-----GRPIVKPEY-----FTEFLKAVQSKKQP 190

Db 181 KASLTLCPEGDWKNKNLDRHNLQDFINIKLNSSSILPEMEGLESETEYLS--ESVEVP 238

QY 191 PQIESFYPP-----LDEP-----STGSKNVDLSGROERKQIFPKGTFTI 228

Db 239 SPFDILEPPTSGGFLKSKPCCYIFPGGRGDSRLFAVNGFNMLINGGSEKSCF----- 292

QY 229 FLNAKHKKLSSAVVFGGE-----ARLTENEHEEH-----NEFLAPGT 268

Db 293 WKLIHRHLDVRDVSILLTHIGDDNLPGINSLMQLRKTALEEEQSQGSTNSDMKNLISPD 352

QY 269 CVVDGTG-TNSQTLIPDCQKKWIOSI-----MDMLQRQGLRPIPEAEIGLAVIFMTTKN 321

Db 353 GVFVFLNYPENLKNMPSFRVK--RSVEENCTFLQYLNKLSMKPEP-----LFRNVGN 402

QY 322 YCDP-----QGHPSSTGLKTTTPGFSLSQ-----GVSYD--EKLMPSA----- 356
 Db 403 TIDPIILFQKMGVGLKEMVLPVSKSKESKVFYHMQWSGTSDKAEFLLPNGQELDIPLS 462
 QY 357 -----PVTNTTVADTESQADTWLSERP----- 381
 Db 463 YFTSVSSLIVHPANPAEKIIRVLFPNGSTQYNILEGLEKHLKHLDFLKQPMVYTKQDLTGN 522
 QY 382 ---KEIKYSKMEQKFRMLSDQAPTVKESCKTSSNNNSMVNTLAKMRIPNYQLSPTKLPS 438
 Db 523 IASPAVKQAKLKQRTDSKESLPAKATTTKQDCQRNKLAKKHWLSQSLVQOQLEKPKOKLS 582
 QY 439 INKS ---KDRASQOOQTNSIRNYFOPSTKK---RERDEENQEMSSCKSARIETSCSLEQT 493
 Db 583 KEKTPVKKEKA-----VRPETKTIIVAKDVTTKBEQLGKS---ETSEKQASEK 627
 QY 494 QPATPSLWKKNKQHLSENPVDTSNNLFTDTDLKSIYKNSASKSHAAEKLSRKNKREM 553
 Db 628 QDVKPKVTKEK---SVKKEVAKPEEK---KDEKEPKKEVSKKEERPLIKKEPKPK 679
 QY 554 DQVATEDEVLEOLFQKDTKPELEIDVKVQKEEDVNVKRPMDIETNDTFSDEAVPESK 613
 Db 680 ED---IKKEVKKEVKEEKEARKEVK-----KEAPPKEAK 712
 QY 614 ISOENEIGKRELKEDSLWSAKEISNNDKLQDDSEMLPKKLLITEPRSLVKNSTSRN-- 671
 Db 713 KEVKKE--EKREIKKEE---KEV-----KKDIKKVPKEIKTPTTPEAKKPAKPKP 760
 QY 672 -----PSGINDDYDGLKNFKPKKVTYPGAGKLPHIIG----- 704
 Db 761 QKKEEPAKKEAVPAGKAKEGKIKTIVKESKVS--GAQTALAAVGAATTAATAAAAEIT 818
 QY 705 --GSDLIAHARKNT-----ELEEWLROEMEVQNHAKESLADD 742
 Db 819 ASGKELEAERSLMSPEDLTQDFEELKABEETIKATRPQVALIED 864
 RESULT 11
 A02986
 myosin alpha heavy chain, cardiac muscle - rabbit (fragment)
 N:Alternate names: alpha isomyosin
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 20-Jun-2000
 C:Accession: A02986; I46865
 R:Kavinsky, C.J.; Umeda, P.K.; Levin, J.E.; Sinha, A.M.; Nigro, J.M.; Jakovcic, S.; Rabi
 J. Biol. Chem. 259, 2775-2781, 1984
 A:Title: Analysis of cloned mRNA sequences encoding subfragment 2 and part of subfragment
 A:Reference number: A02986; MUID:84135762
 A:Accession: A02986
 A:Molecule type: mRNA
 A:Residues: 1-465 <KAV>
 A:Note: the authors translated the codon AAG for residue 149 as Arg
 R:Sinha, A.M.; Umeda, P.K.; Kavinsky, C.J.; Rajamanickam, C.; Hsu, H.
 Proc. Natl. Acad. Sci. U.S.A. 79, 5847-5851, 1982
 A:Title: Molecular cloning of mRNA sequences for cardiac alpha- and beta-form myosin hea
 A:Reference number: I46865; MUID:83299886
 A:Accession: I46865
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 233-270, 'L', 272-353 <SIN>
 A:Cross-references: GB:J00671; NID:G165528; PIDN:AAA31410.1; PID:G165529
 C:Comment: In the rabbit the expression of alpha and beta isomyosins follows a defined o
 n can be altered by thyroid hormone as well as by other stimuli.
 C:Comment: Protease-sensitive junctional regions located at the head-rod (SI-S2 subfragm
 sin head and actin filament during contraction.
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; cardiac muscle; coiled coil; heart; muscle; thick filament
 F:1-36/Domain: head (fragment) <CAR>
 F:37-465/Domain: rod (fragment) <ROD>

Query Match 4.4%; Score 171; DB 2; Length 465;
 Best Local Similarity 21.3%; Pred. No. 0.057;
 Matches 81; Conservative 67; Mismatches 136; Indels 96; Gaps 14;
 QY 384 IKVSKMEQKFRMLSDQAPTVKESCKTSSNNNSMVNTLAKMRIPNYQLS---PTKLPSIN 440
 Db 37 LKSAEAEKEMAAMKEEFGRIKESLEKSEARKELEKEMVSLLOEKNDLQLOVQAEQDNLN 96
 QY 441 KSKDRASQOOQTNSIRNYFOPSTKK---ERDEENQEMS---SCKSARIETSCS--LLEQT 493
 Db 97 DAERCDQ-----LIKKIQLEAKVKEMNERLEDEEMNAELAKRKLDECESELKKDI 151
 QY 494 QPATPSLWK--NKEQHLSENPVDTSNNLFTDTDLKSIYKNSASKSHAAEKLSRKNK-- 550
 Db 152 DDLELTAKVEKEKHATENK-----VKNLTEEMAGLDEIIAKTLKKEKAL 196
 QY 551 -----REMDVATEDEVLEOLFQKDTKPELEIDVKVQKEEDVNVKRPMDIETNDTFS 605
 Db 197 QEAHQALDDQLQAEDEKVNLT---TKAKLKEQOVDLLEGSLEQEKVKRMDLER----- 247
 QY 606 EAVPESKISQENEIGKRELKEDSLWSAKEISNNDKLQDDSEMLPKKLLITEPRSLV 665
 Db 248 -----AKRLEGDLKTQESIND---LENDKQOLEERLKKKEF----- 282
 QY 666 NSTSRNPSSGINDDYDGLKNFKPKKVTYPGAGKLPHIIGSDLIHARKNTELEEWLRQ 725
 Db 283 -DISQLNSKIEDQALVLQOKKLK-----ENQARIEELEEELEA 321
 QY 726 EMEVQNHAKESLADDLFR 745
 Db 322 E---RTARAKVEKLRSDLSR 338
 RESULT 12
 I48175
 myosin heavy chain alpha, cardiac muscle [similarity] - golden hamster
 C:Species: Mesocricetus auratus (golden hamster)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Feb-2001
 C:Accession: I48175; A23938
 R:Wang, R.; Sole, M.J.; Cukerman, E.; Liew, C.C.
 J. Mol. Cell. Cardiol. 26, 1155-1165, 1994
 A:Title: Characterization and nucleotide sequence of the cardiac alpha-myosin heavy c
 A:Reference number: I48153; MUID:95115033
 A:Accession: I48175
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1939 <RES>
 A:Cross-references: GB:I15351; NID:9402373; PIDN:AAB59701.1; PID:9402374
 R:Liew, C.C.; Jandreski, M.A.
 Proc. Natl. Acad. Sci. U.S.A. 83, 3175-3179, 1986
 A:Title: Construction and characterization of the alpha-form of a cardiac myosin hea
 A:Reference number: A23938; MUID:86205859
 A:Accession: A23938
 A:Molecule type: mRNA
 A:Residues: 1630-1843, 'R', 1845-1878, 'T', 1880-1927, 'N', 1929-1932, 1934-1939 <LIE>
 C:Genetics:
 A:Introns: 67/3; 115/3; 168/1; 177/2; 214/3; 245/3; 267/1; 300/1; 334/3; 381/1; 470/3
 392/2; 1453/3; 1509/1; 1550/3; 1653/3; 1721/3; 1763/3; 1855/3; 1887/3; 1932/3
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: ATP; cardiac muscle; heart; nucleotide binding; P-loop
 F:88-768/Domain: myosin motor domain homology <MMOT>
 F:178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 4.4%; Score 171; DB 2; Length 1939;
 Best Local Similarity 23.2%; Pred. No. 0.39;
 Matches 89; Conservative 61; Mismatches 130; Indels 104; Gaps 17;
 QY 384 IKVSKMEQKFRMLSDQAPTVKESCKTSSNNNSMVNTLAKMRIPNYQLS---PTKLPSIN 440
 Db 842 LKSAETEKEMANKKEEFGRVKESLEKSEARKELEKEMVSLLOEKNDLQLOVQAEQDNLN 901
 QY 441 KSKDRASQOOQTNSIRNYFOPSTKKRE-----RDEE--NOEMSSCKSARIETSCS--LLEQ 492

```

Db 902 DAERCQ-----LKKKIOLEAKVKEMTERLEDEEEMNAELTS-KARKLEDECESEUKD 955
QY 493 TOPATPSLWK-NKEQHLSENPEVDTSNNDNLFTDILKSIVKNSASKASHAAKLSRNNK- 550
Db 956 IDLELTAKVERKHATENK-----VKNLTEEMAGLEIIAKLTKEKA 1000
QY 551 -----REMDVAIEDVLBOLFKDKPELEIDVVKQKQBEDVNVRRPRMDIETNDTFS 604
Db 1001 LOEAHQALDLOAEEDKVNTL---TKSKVKLEQVDDLEGSLEQKKVRMDLER----- 1052
QY 605 DEAVPESKISQENIGKKEKRELKEDSLWSAKEISN--NDKLQDDSEMLPKLLITPEPSL 662
Db 1053 -----AKRKLGDLVNTQESIMDLENDKIQ-----LEELKKKEF--- 1087
QY 663 VIKNSTSRNPSSGINDDYG-OLKNFKFKPVYTPGAGKLPHIIGSGDLIAHARKNTELEE 721
Db 1088 ----DISQONSKIEDQALALQKKLKE-----NQARIELEE 1122
QY 722 WLQRMENVQNHAKESLADDLFR 745
Db 1123 ELEAE---RTARAKVEKLRSDLTR 1143

```

RESULT 13

ORMSPI

```

microtubule-associated protein MAP1B - mouse
N:Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protein
C:Species: Mus musculus (house mouse)
C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 01-Sep-2000
C:Accession: S07549; S44387; A33645
R:Noble, M.; Lewis, S.A.; Cowan, N.J.
J. Cell Biol. 109, 3367-3376, 1989
A:Title: The microtubule binding domain of microtubule-associated protein MAP1B contains
A:Reference number: A33645; MUID:90094539
A:Accession: S07549
A:Molecule type: mRNA
A:Residues: 1-2464 <NO>
A:Cross-references: EMBL:X51396; NID:g52999; PIDN:CAA35761.1; PID:g53000
R:Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.
Arch. Biochem. Biophys. 310, 428-432, 1994
A:Title: Binding of heat-shock protein 70 (hsp70) to tubulin.
A:Reference number: S44387; MUID:94234720
A:Accession: S44387
A>Status: preliminary
A:Molecule type: protein
A:Residues: 653-663, 'IC' <SAN>
C:Superfamily: microtubule-associated protein MAP1B
C:Keywords: microtubule binding; phosphoprotein; tandem repeat
F:589-786/Domain: microtubule binding #status experimental <MTB>
F:589-592,639-642,649-652,655-658,660-663,668-671,674-677,679-682,683-686,687-690,691-69
R-K-E/D-X)
F:1861-2064/Region: 17-residue repeats
F:91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/binding site: ph
F:147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (co
F:1953/Binding site: phosphate (Tyr) (covalent) #status predicted

```

```

Query Match 4.4%; Score 170; DB 1; Length 2464;
Best Local Similarity 18.0%; Pred. No 0.6;
Matches 171; Conservative 148; Mismatches 301; Indels 330; Gaps 44;

```

```

QY 6 PAAGPAGGEPYRLLTGVYVGRKNCAILIENDQISRNH---AVLTANFSV-----TNLS 58
Db 20 PAASGPSLSHRLDSKFY-----LLVVGVETVEEHLRAIGNIELGIRSDNLI 71
QY 59 QTD---EIPVLTKDKNKYGTFVNEEKMQNGFSRTLK-----SGDGITFGVFGSKFRI 108
Db 72 ECNLDQELKLFVSRHSARFSPVPGOKILHRSDVLETVVVLINPSDEAVS-----TEVRL 126
QY 109 -----EYELPVACSSCLDYSKLTALNOAILQLGGFTVNNWTECHTLVNVSVKVI--- 159
Db 127 MITDAARHKLVLITGQCFENFG-----ELILQSGSFSFQNFIEFTDQEIPELLSTHPA 181

```

```

QY 160 -KTICALIC-----GRPIVKPEY-----FTEFLKAVQSKKQ 189
Db 182 NKASLTLCFCEEGDWKNSNDRHNLQDFINIKLNSASILPEMEGLSEFTEYLS--ESVEV 239
QY 190 PPOIESYPP-----LDPE-----SIGSKNVDLSGRQERKQIFKGTFF 227
Db 240 PSPFDILEPTSGGFLKLSKPCCYIFPGRGDSALFVNGFNMLINGSERKSCF----- 294
QY 228 IFLNAKHKKLSAVVFGGE-----ARLITEENEENHFFLAPGTCVVVDITGTTN 277
Db 295 -WKLIRHLDRVDSILLTHIGDDNLPGINSMLQKIAELEER-----SQGSTS 341
QY 278 SQTLLPDCQKKWISIMDLQROGLRPIPEAEIGLAVIFMTT-KNYCDPQGHPTGLKTT 336
Db 342 N-----SDMKNLIS-----PD-----LGVVFLNVPENLKDPE----- 369
QY 337 TPGPSLSQGSVDKELMPSAPVNTTIVYADTSEQADTWLSEPK----- 382
Db 370 ---PNTKMKRSIEACFTLOYLN-----KLSMKPEPLFRSVGNWTIEPVI 410
QY 383 -----EIKVSKMEQKFRMLSQDAPTVKESCKTSSNNNSMV----- 417
Db 411 LFOKMGVGLKLEMYVLNPNVSKSEMQYF--MQOWTGTNKKRAELLIPNGQEVDDIPISYLS 468
QY 418 -----SNTAKMRIPNYQLSPTKLPISINKS-----DRASQOQOTNSI--- 455
Db 469 VSSLIVWHPANPAEKIIRVLFPNGSTQYNILEGLEKHLKHLDFLKQPLATOKDLTGQVPTP 528
QY 456 -----RNYPOPSTK-----KREDEENQEMSSKSARLETSCSLEQTQP 495
Db 529 PVQVKLIKQRADRESLSUKPATKPVASKVRKSEKTEPVT--KTSQVE----- 575
QY 496 ATPSLMKNKQHLSENPEVDTSNNDNLFTDILKSIVKNSASKASHAAKLSRNNKREM-D 554
Db 576 KTPKVESKEKVLVKKDKPVATESKPSV-TEKEYSSKEEQSPVKAEEVAEKATESKPKVTK 634
QY 555 DVAIEDSV---LEQ-----LFDKTPK---ELEIDVKVQ-KQEDVNVV 589
Db 635 DKVVKKEIKTKLEKKEKPKKEVVKKEDKTPPLKDKPKRKEEVKKEIKKEERKEL 694
QY 590 RKRPRMDIETNDTFSDEAVPESKISQENIGKKEKRELKEDSLWSAKEISNNDKLQD--D 646
Db 695 KKEVKKETPLKDA-KKEVKKEEKKEKKEKPKKEIKKIS-----KDKIKSTPOSDDTKKP 749
QY 647 SEMLPKLLLTFRSLVIKNSTSRNP--SGINDDYGOLKNFKFKKVTYPG-----AGK 698
Db 750 SALPKPVAKKEE-----STKKEPLAAGKLKDKGKVKVKKKEKTEAAATAVGTAAT 801
QY 699 LPHIIGSGDLIAHARKNTELEEWLRQEME-----VQNHAKESLADDL 743
Db 802 TAAVVAAGAAAGSPVKVELEASRLMSSPEDLTAKDFEELKAEEDVAKDI 851

```

RESULT 14

S20117

```

protein kinase BCK1 (EC 2.7.1.1) - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein J0906; protein kinase SLK1; protein kinase SSP31; protein
C:Species: Saccharomyces cerevisiae
C>Date: 23-Apr-1993 #sequence_revision 23-Apr-1993 #text_change 24-Sep-1999
C:Accession: S20117; S50298; S22285; S19061; JQ1432; S56872; S30794; JQ1118
Mol. Cell. Biol. 12, 1162-1178, 1992
R:Costigan, C.; Gehring, S.; Snyder, M.
A:Title: A synthetic lethal screen identifies SLK1, a novel protein kinase homolog im
A:Reference number: S20117; MUID:92186847
A:Accession: S20117
A:Molecule type: DNA
A:Residues: 1-1478 <COS>
A:Cross-references: EMBL:M84389
A:Experimental source: strain S288C
R:Miosga, T.; Boles, E.; Schaaff-Gerstenschlaeger, I.; Schmitt, S.; Zimmermann, F.K.
Yeast 10, 1481-1488, 1994
A:Title: Sequence and function analysis of a 9.74 kb fragment of Saccharomyces cerevi

```

```

Qy 502 KNKEOHL-----SNEPVDVTNSNNLFTDIDLKSLVKNKSAKSHAAEKL 545
| : : : | | | : | | | : | | | : | | | : | | | : | | |
Db 532 KYSVNFFLDQOKFYPLKKTGLTGLNDSENKYLTVTKDNVSVFVPLNLSVAKLSFKEASALTKL 591
| : : : | | | : | | | : | | | : | | | : | | | : | | |
Qy 546 RSNKKR---EMDD-----VALEDEVLE---OLFKDTKPELEI--DVKVOKO----- 583
| : : : | | | : | | | : | | | : | | | : | | | : | | |
Db 592 GINHKNVTFHMTDFDCDIGAALPDOTLEFLKKSLELNTSGKIYIKDMKLQOKRPKAPLPT 651
| : : : | | | : | | | : | | | : | | | : | | | : | | |
Qy 584 -EEDY---NVKRPRMDIETNDFSD---EAVPESSKISQENE----- 619
| : : : | | | : | | | : | | | : | | | : | | | : | | |
Db 652 SENNVPLAKSVKSKSMRSGTSLIASTDVSVITSSSDITSPDEHAGSGRRYPQTPSY 711
| : : : | | | : | | | : | | | : | | | : | | | : | | |
Qy 620 ---ICKRELKEDSLWSAKEISNNDKLODSEMLPKKLLLTFR---SLVKNKSTSRNPS 673
| : : : | | | : | | | : | | | : | | | : | | | : | | |
Db 712 YDRVSTNTPTBELNWNVIKEVLSHE-----ENAPKMVFKTPKLELNLDPKGSKLNIPT 765
| : : : | | | : | | | : | | | : | | | : | | | : | | |
Qy 674 GINDYTGQLKFKPKFKKVTYPGAGKLPHLIIGSDLIAHHARKN 716
| : : : | | | : | | | : | | | : | | | : | | | : | | |
Db 766 PITNESK-SSFQVLURKDE-----GTEIDFNHRRES 795
| : : : | | | : | | | : | | | : | | | : | | | : | | |

RESULT 15
T21312
hypothetical protein F23D12.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T21312; T24907
R:Barlow, K.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19404
A:Accession: T21312
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-2287 <W1>
A:Cross-references: EMBL:Z71186; PIDN:CAA94917.1; GSPDB:GN000028; CESP:F23D12.2
A:Experimental source: clone F23D12
R:Barlow, K.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19952
A:Accession: T24907
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-2287 <W12>
A:Cross-references: EMBL:Z70687; NID:g1256502; PIDN:CAA94618.1; GSPDB:GN000028;
A:Experimental source: clone T14C1
C:Genetics:
A:Gene: CESP:F23D12.2
A:Map position: X
A:Introns: 75/3; 141/3; 165/3; 409/3; 497/3; 572/1; 583/3; 686/3; 773/3; 916/3

```

[illegible]

```

Db 1321 AVVSNGLGQYHH-----MSMSLTEAE--VFSCNR-----1347
QY 301 GLRPIPAEIGLAVIFMTTKNYCD-----POGHPSTGLKTTTP--GPSLSQGVSVDEK 351
Db 1348 -----FOGRISYCSDMWEYKPOS--LEALNPTFPKIGTCTPSSSTVSMADN 1392
QY 352 LMPSPAVNTTYYVADTESEQADTWLSERPKEIKVSKMEQFRMLSDQAPTAVKESCKTSS 411
Db 1393 LSGILELHSSFDLTNTTSKSVNTPPFGEYPTMI-----WEAPKL---CTSDN 1436
QY 412 NNSWSVN-----TLAK--MRIPNYQLSTPKLPSPINKSKDRASQQOQTN---453
Db 1437 NNNNSCKNDQORNEEKVDKTRSLIKLSKIANVYFGPFPSAITKNSNFAIKPLEINEQP 1496
QY 454 SIRNYFOPSTK-----KREDEENQEM-----SSCKSARIETSCSLEOTQOPATPSL 500
Db 1497 DIINCNPVSKSAVSKTNLKNVSNYKQKMHFAKSHFHPADLEMPADLEISLMRT---1553
QY 501 WKNKEQHLSNEPVDT--NSDNNLFTDTDLKSIYKNSASKSHAAEKLRSNKREMDDVAI 558
Db 1554 -NNTESNLKNTKTIATISINEMDMISDDNYPMLQNNVAK-----HQKRKVEICSL 1603
QY 559 EDEVLEQLFKDTPELEIDVKVQOEEDVNVKRPRMDIETNDFSD-----605
Db 1604 DNNVYTSRKLTKRELLKCS-KTEKPDYEPSSAKTPQGNFFKGDSSAKIDLFKGSHKFIKD 1662
QY 606 ---EAVPESSKISOENEIGKRELKEDSL-----WSAKEI-----637
Db 1663 ERAETCPDKTEIPEEVIQSTQNLHENKAIKDKVVPINANNLKQKNVKEIPTTVDVGQLP 1722
QY 638 -----SNNDKLODDSEMLPKLL-LTEFR-----SLVIKNSTSRNPSPGINDDYG-----680
Db 1723 ILDYNYSNRSRTNENLFSKVKFGKNEYQYGPSMGKLIERMGTGTGPKVYVSEYSSSPIA 1782
QY 681 -----OLKNFKFKKVYTP-----GAGKLPHIIGGSDLIAHHARKNTELEEWL 723
Db 1783 AIFSPKADSHEFADSNKLOKTTKPSICMKFGNCGRLPDL-----EPNKN-----WL 1829
QY 724 ROEM-EVQNOHA-----KEESLADDLF 744
Db 1830 RONNAQFAGHHASVIDTSKSTFKLKTSLVQWF 1863

```

Search completed: August 15, 2002, 01:42:27
 Job time: 5262 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: August 15, 2002, 01:41:25 ; Search time 24.79 Seconds
(without alignments)
1177.674 Million cell updates/sec

Title: US-09-837-602-2
Perfect score: 3899
Sequence: 1 MWKLLPAAGPAGGEPYRLLT.....KEPSLADLFRYNPYLKRRR 754

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	179.5	4.6	1790	1 US01_YEAST	P25386 saccharomyc
2	173	4.4	1938	1 MYH6_MOUSE	Q02566 mus musculu
3	171	4.4	465	1 MYH6_RABIT	P04460 oryctolagus
4	171	4.4	1939	1 MYH6_MESAU	P13539 mesocricetu
5	170.5	4.4	1939	1 MYH6_HUMAN	P13533 homo sapien
6	170.5	4.4	2492	1 ATRX_HUMAN	P46100 homo sapien
7	170	4.4	2464	1 MAPB_MOUSE	P14873 mus musculu
8	169.5	4.3	1478	1 BCK1_YEAST	Q01389 saccharomyc
9	168.5	4.3	1233	1 YF16_YEAST	P43597 saccharomyc
10	168.5	4.3	3672	1 LML2_CAEEL	Q21313 caenorhabdi
11	167.5	4.3	2349	1 TPR_HUMAN	P12270 homo sapien
12	166	4.3	1359	1 ATRX_CAEEL	Q9u7e0 caenorhabdi
13	166	4.3	1935	1 MYH7_HUMAN	P12883 homo sapien
14	165	4.2	667	1 CYLL_BOVIN	P35662 bos taurus
15	165	4.2	1938	1 MYH6_RAT	P02563 rattus norv
16	165	4.2	1939	1 MYH1_HUMAN	P12882 homo sapien
17	164.5	4.2	2871	1 DESP_HUMAN	P15924 homo sapien
18	164	4.2	915	1 ZDS1_YEAST	P50111 saccharomyc
19	164	4.2	1381	1 YBE7_YEAST	P34216 saccharomyc
20	163	4.2	952	1 YK15_CAEEL	P46012 caenorhabdi
21	162.5	4.2	539	1 YAH2_YEAST	P39705 saccharomyc
22	162.5	4.2	1427	1 REST_HUMAN	P30622 homo sapien
23	162.5	4.2	1957	1 YDB6_SCHPO	Q10411 schizosacch
24	162	4.2	1312	1 RA50_YEAST	P12753 saccharomyc
25	162	4.2	2748	1 NUM1_YEAST	Q00402 saccharomyc
26	161	4.1	1928	1 MYSL_YEAST	P08964 saccharomyc
27	160	4.1	1934	1 MYH7_MESAU	P13540 mesocricetu
28	159.5	4.1	1189	1 YJH6_YEAST	P47035 saccharomyc
29	158.5	4.1	1664	1 INT1_CANAL	P53705 candida alb
30	158	4.1	1744	1 TANA_XENLA	Q01550 xenopus lae
31	158	4.1	1938	1 MYH4_RABIT	Q28641 oryctolagus
32	157.5	4.0	1937	1 MYH8_HUMAN	P13535 homo sapien
33	157	4.0	1164	1 BAG_STRAG	P27951 streptococc

RESULT 1									
ID	US01_YEAST	STANDARD;	PRT;	1790	AA.				
AC	P25386;								
DT	01-MAY-1992 (Rel. 22, Created)								
DT	01-MAY-1992 (Rel. 22, Last sequence update)								
DT	16-OCT-2001 (Rel. 40, Last annotation update)								
DE	Intracellular protein transport protein US01.								
GN	US01 OR INT1 OR YDL058W.								
OS	Saccharomyces cerevisiae (Baker's yeast).								
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;								
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.								
OX	NCBI_TaxID=4932;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=X2180-1A;								
RX	MEDLINE=91185402; PubMed=2010462;								
RA	Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,								
RA	Yamasaki M.;								
RT	"A cytoskeleton-related gene, usol, is required for intracellular								
RT	protein transport in Saccharomyces cerevisiae.";								
RL	J. Cell Biol. 113:245-260(1991).								
RN	[2]								
RP	SEQUENCE OF 782-1790 FROM N.A.								
RA	Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,								
RA	Kendrick K.E.;								
RL	Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.								
[3]									
RP	SEQUENCE OF 1-8 FROM N.A.								
RA	Bai Y., Symington L.S.;								
RL	Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.								
CC	-!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI								
CC	COMPLEX.								
CC	-!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR								
CC	MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE								
CC	ER AND THE GOLGI COMPLEX.								
CC	-!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED								
CC	OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL								
CC	COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.								
CC	-!- SIMILARITY: BELONGS TO THE VDP/US01/YBL047C FAMILY.								
CC	-----								
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration								
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -								
CC	the European Bioinformatics Institute. There are no restrictions on its								
CC	use by non-profit institutions as long as its content is in no way								
CC	modified and this statement is not removed. Usage by and for commercial								
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/								
CC	or send an email to license@isb-sib.ch).								
CC	-----								
DR	EMBL; X54378; CAA38253.1; -								
DR	EMBL; L03188; AAB00143.1; -								
DR	EMBL; U53668; AAB66659.1; -								
DR	PIR; A38455; A38455.								
DR	HSSP; P80220; LDIP.								
DR	SGD; S0002216; US01.								
DR	InterPro; IPR002017; Spectrin.								

34	156.5	4.0	671	1 CHS5_YEAST	Q12114 saccharomyc
35	156.5	4.0	783	1 YAYB_SCHPO	Q10218 schizosacch
36	156.5	4.0	1630	1 MSP1_PLAFK	P04932 plasmodium
37	156.5	4.0	1639	1 MSP1_PLAFW	P04933 plasmodium
38	156	4.0	598	1 CYL1_HUMAN	P35663 homo sapien
39	156	4.0	736	1 MYH7_RABIT	P04461 oryctolagus
40	156	4.0	1084	1 MYSS_RABIT	P02562 oryctolagus
41	155.5	4.0	976	1 SCP1_HUMAN	Q15431 homo sapien
42	155.5	4.0	1325	1 YAB6_SCHPO	Q09847 schizosacch
43	155.5	4.0	1941	1 MYH2_HUMAN	Q9ukx2 homo sapien
44	155.5	4.0	2663	1 CENE_HUMAN	Q02224 homo sapien
45	154.5	4.0	1526	1 MYB2_SCHPO	Q9usi6 schizosacch

ALIGNMENTS

KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
 FT DOMAIN 1 724 GLOBULAR HEAD.
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
 FT CONFLICT 847 847 G -> E (IN REF. 2).
 FT CONFLICT 924 924 E -> K (IN REF. 2).
 FT CONFLICT 1253 1253 V -> I (IN REF. 2).
 FT CONFLICT 1319 1319 I -> V (IN REF. 2).
 FT CONFLICT 1461 1461 N -> S (IN REF. 2).
 FT CONFLICT 1581 1581 G -> S (IN REF. 2).
 FT CONFLICT 1600 1600 I -> V (IN REF. 2).
 FT CONFLICT 1661 1661 R -> S (IN REF. 2).
 FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
 SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B21659FD4818 CRC64;

 Query Match 4.6%; Score 179.5; DB 1; Length 1790;
 Best Local Similarity 23.1%; Pred. No. 0.065;
 Matches 100; Conservative 77; Mismatches 168; Indels 87; Gaps 19;

 QY 378 SERPKETKYSKQKFMLSODAPTCKTSNNNSMVSNTLAKMRIPYQLSPKLP 437
 Db 988 SKNESSQLSNLQNKIDSMQEKENFQIERGSIKIEQLKKTIS-----DLEQTKEE 1040
 QY 438 SINK---SKDRASQ-----OQTNSIRNYFOPSTKKRDERDEBNQMSCKSARIE 484
 Db 1041 IISKSDSSKDEYESQISLLKEKLETATTANDENVNKISLTKTREETLEAEALAYKNLNE 1100
 QY 485 TSCSLLEQTOPATPSLWKNKEQHL-----ENEPVDNTSDNNLFTDTDLKSIK 535
 Db 1101 LE-TKLETSEKALKEV-KENEELKKEKIQLEKEATETKQOLNSLR-ANLESEKEHEDL 1157
 QY 536 ASKSHAAKLRNKKREMD--VAIEDEVL-----EQLFKDTKPELEIDVKKVK--QEED 586
 Db 1158 AALQKVEEQIANKOYNEIEISQNLDEITSTQOENESIKKNDLEGEVKAAMKTSSEQ 1217
 QY 587 VNYKRP-----RMDIETNDTDFDEAVPESKISQENEIGKREKLEDKEDLSWSAK 635
 Db 1218 SNLKSEIDALNLQIKELKKKNETNEA-----SLESIK-SVESETVKIKELQDECNFKEK 1272
 QY 636 EISN-NDKLQDDSEMLPKLLLETFSLVTKNSTSRNPISINDYQOLKNFKFKKVTYP 694
 Db 1273 EVSELEDKLR-ASEDKNSKYLEQKSEKIEKELDAKTTKLTQLEKITNLRSKAKESSES 1331
 QY 695 GAGKLPHIIGGSDLIHARKNTELE-EWLRQEMEVQNHAKEE----- 737
 Db 1332 ELSRLKK-----TSSEERKNAEEQLEKLNQIKNQAFERKLLNBSGTSITQEYS 1384
 QY 738 ----SLADLFR 745
 Db 1385 EKINTLEDELIR 1396

 RESULT 2
 MYH6_MOUSE STANDARD; PRT; 1938 AA.
 AC Q02566; Q64258; Q64738;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain, cardiac muscle alpha isoform (MYHC-alpha).
 GN MYH6 OR MYHCA.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C, A/J, C57BL/6J, AND DBA/2J;
 RX MEDLINE=92250040; PubMed=1577481;
 RA Quinn-Laquer B.K., Kennedy J.E., Wei S.J., Beisel K.W.;

RT "Characterization of the allelic differences in the mouse cardiac
 alpha-myosin heavy chain coding sequence.";
 RL Genomics 13:176-188(1992).
 RN [2]
 RP SEQUENCE OF 1-67 FROM N.A.
 RC STRAIN=AKR;
 RX MEDLINE=91225025; PubMed=2026617;
 RA Gulick J., Subramanian A., Neumann J., Robbins J.;
 RT "Isolation and characterization of the mouse cardiac myosin heavy
 chain genes";
 RL J. Biol. Chem. 266:9180-9185(1991).
 CC -!- FUNCTION: MUSCLE CONTRACTION.
 CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE. SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
 MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M76598; AAA371159.1; -;
 DR EMBL: M76599; AAA371160.1; -;
 DR EMBL: M76600; AAA371161.1; -;
 DR EMBL: M76601; AAA371162.1; -;
 DR EMBL: M62404; AAA37424.1; -;
 DR HSSP: P08799; LMMD.
 DR SWISS-2DPAGE: Q02566; MOUSE.
 DR MGD: MGI:97255; Mvha.
 DR InterPro: IPR000048; IQ.
 DR InterPro: IPR004009; Myosin_N.
 DR InterPro: IPR002928; Myosin_tail.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00612; IQ; 2.
 DR Pfam: PF00063; myosin_head; 1.
 DR Pfam: PF02736; Myosin_N; 1.
 DR Pfam: PF01576; Myosin_tail; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR ProDom: PD000355; myosin_head; 1.
 DR SMART: SM00015; IQ; 1.
 DR SMART: SM00242; MYSC; 1.
 DR PROSITE: PS50096; IQ; 1.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family;
 KW Calmodulin-binding; Polymorphism.
 FT DOMAIN 1 782 MYOSIN HEAD-LIKE.
 FT DOMAIN 783 812 IQ.
 FT DOMAIN 812 1938 COILED COIL (POTENTIAL).
 FT NP_BIND 178 185 ATP.
 FT DOMAIN 657 679 ACTIN-BINDING.
 FT DOMAIN 759 790 ACTIN-BINDING.
 FT DOMAIN 816 833 CALMODULIN-BINDING (BY SIMILARITY).
 FT DOMAIN 129 129 METHYLATION (TRI-) (POTENTIAL).
 FT MOD_RES 697 697 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 707 707 ALKYLATION (SH-2) (POTENTIAL).

```
FT VARIANT 194 194 Y -> D.
FT VARIANT 545 838 S -> A.
FT VARIANT 838 838 I -> S.
SQ SEQUENCE 1938 AA; 223564 MW; EAD789ADA68818FB CRC64;

Query Match 4.4%; Score 173; DB 1; Length 1938;
Best Local Similarity 22.7%; Pred. No. 0.16;
Matches 92; Conservative 61; Mismatches 122; Indels 130; Gaps 17;

QY 365 ADTESEADTWLSEPK-----ELKVSMEQKFRMLSDQAPTVKESCKTSNNNSMWS 418
   :||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 845 AETEKEMANKEEFGVKDALEKSEARRKELEKMSLLQEKNDLQVQAEQDN----- 899
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 419 NTLAKMRIPYQLSPKLPINKSKRASQOQOQTSIRNYFQPSFKKR-----ERDEENQEM 475
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 900 -----LNDAEERCQ-----LIKNKIQLQLEAKVKEMTERLEDEEM 934
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 476 S-----SCKSARIETSCS-LLEQTOPATPSLWK-NKEOHLSENEPVDNNDNLFDTDLKS 530
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 935 NAELETAKKRLEDECSKELKDDIDLETLAKVEKEKHATENK-----VKN 979
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 531 IVANSASKSHAAKLSNKK-----REMDVVAIEDVLEQLFKDTPKELEIDVKVQKQ 583
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 980 LTEEMAGLDLIIAKLTKEKKALQEAHQALDDQAEEDKVNTL---TKSVKLEQQVDDL 1036
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 584 EEDVNVKRPMDIETNDTFSDFAVPESKISQENETGKKRELKEDSLWSAKEISN--ND 641
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1037 EGSLQEKVKRMDLER-----AKRLEGDLKLTQESIMDLEND 1074
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 642 KLQDDSEMLPKLLLTFFRSVLVKNSTSRNPSPGINDYQ-OLKNFKFKFKVTPYPGAKLP 700
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1075 KLQ-----LEELKKKEF-----DISQONSKIEDQALALQLOKKLKE----- 1112
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 701 HIIGSDLIHAHARKNTELEWLRQMEVONQHAKESLADDLFR 745
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1113 -----NQARTEEELEEA-----RTARAKVEKLSRDLR 1143
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
MYH6_RABIT STANDARD; PRT; 465 AA.
AC P04460;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle alpha isoform (MyHC-alpha) (Alpha
DE isomyosin) (fragment).
GN MYH6.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid:9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84135762; PubMed=6321481;
RA Kavinsky C.J., Umeda P.K., Levin J.E., Sinha A.M., Nigro J.M.,
RA Jakovcic S., Rabinowitz M.;
RT "Analysis of cloned mRNA sequences encoding subfragment 2 and part of
RT subfragment 1 of alpha- and beta-myosin heavy chains of rabbit
RT heart.";
RL J. Biol. Chem. 259:2775-2781(1984).
RN [2]
RP SEQUENCE OF 91-177 FROM N.A.
RX MEDLINE=84221901; PubMed=6328491;
RA Friedman D.J., Umeda P.K., Sinha A.M., Hsu H.J., Jokovic S.,
RA Rabinowitz M.;
RT "Characterization of genomic clones specifying rabbit alpha- and
RT beta-ventricular myosin heavy chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:3044-3048(1984).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
```

```
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; K02443; AAA31412.1; -.
CC EMBL; K01697; AAA31416.1; -.
CC DR PTR; A02986; A02986.
CC DR HSP; P13538; 2MYS.
CC DR InterPro; IPR002928; Myosin_tail.
CC DR Pfam; PF01576; Myosin_tail; 1.
CC KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC ATP-binding; Multigene family.
CC FT NON_TER 1 1 MYOSIN HEAD-LIKE.
CC FT DOMAIN 36 >465 COILED COIL (POTENTIAL).
CC FT NON_TER 465 465
CC SQ SEQUENCE 465 AA; 54375 MW; BAC86AFF8CE8E29C CRC64;

Query Match 4.4%; Score 171; DB 1; Length 465;
Best Local Similarity 21.3%; Pred. No. 0.033;
Matches 81; Conservative 67; Mismatches 136; Indels 96; Gaps 14;

QY 384 IKVSKMEQKFRMLSDQAPTVKESCKTSNNNSMVSNTLAKMRIPNYQLS---PTKLPSIN 440
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 37 LKSAEAKEMAAKKEEFGRIKESLEKSEARRKELEKMSVLLQEKNDLQVQAEQDNLN 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 441 KSKDRASQOQOOTSIRNYFQPSFKKR---ERDEENQEMS---SCKSARIETSCS-LLEQT 493
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 97 DAEERCQ-----LIKNKIQLQLEAKVKEMNERLEDEEMNAELTAKRKLDECSKELKDI 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 494 QPATPSLWK-NKEOHLSENEPVDNNDNLFDTDLKSIKNSASKSHAAKLSNKK-- 550
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 152 DDLTLAKVEKEKHATENK-----VKNLTEEMAGLDLIIAKLTKEKKAL 196
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 551 -----REMDVVAIEDVLEQLFKDTPKELEIDVKVQOEEDVNVKRPMDIETNDTFSD 605
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 197 QEAHQALDDLQAEEDKVNTL---TKAKLLEQQVDDLEGSLEQEKVKRMDLER----- 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 606 EAVPESSKISQENETGKKRELKEDSLWSAKEISNNKLDQDSEMLPKLLLTFFRSVLK 665
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 248 -----AKRLEGDLKLTQESIMD---LENDKQOLEERLKKKEF----- 282
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 666 NSTSRNPSPGINDYQGLKNFKFKFKVTPYPGAKLP HIIGSDLIHAHARKNTELEWLRQ 725
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 283 -DISQONSKIEDQALVLOQKKLKE-----ENQARTEEELEEA 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 726 EMEVONQHAKESLADDLFR 745
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 322 E---RTARAKVEKLSRDLR 338
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
MYH6_MESAU STANDARD; PRT; 1939 AA.
ID MYH6_MESAU
AC P13539; Q06562;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
```

16-OCT-2001 (Rel. 40, Last annotation update)
 MYH6.
 Mesocricetus auratus (Golden hamster).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 Mesocricetus.
 NCBI_TaxID=10036;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=FJB; TISSUE=Liver;
 MEDLINE=95115033; PubMed=7815459;
 Wang R., Sole M.J., Cukerman E., Liew C.-C.;
 "Characterization and nucleotide sequence of the cardiac alpha-myosin
 heavy chain gene from Syrian hamster";
 J. Mol. Cell. Cardiol. 26:1155-1165(1994).
 [2]
 SEQUENCE OF 1630-1939 FROM N.A.
 MEDLINE=86205859; PubMed=3458174;
 Liew C.-C., Jandreski M.A.;
 "Construction and characterization of the alpha form of a cardiac
 myosin heavy chain cDNA clone and its developmental expression in the
 Syrian hamster";
 Proc. Natl. Acad. Sci. U.S.A. 83:3175-3179(1986).
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE. SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY (BY
 SIMILARITY).
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC -1- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
 MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L15351; AAB59701.1; -;
 CC EMBL: M12995; AAA37081.1; -;
 CC HSP: P08799; IMMD.
 CC InterPro: IPR000048; IQ.
 CC InterPro: IPR004009; Myosin_N.
 CC InterPro: IPR002928; Myosin_tail.
 CC InterPro: IPR001609; myosin_head.
 CC Pfam: PF00612; IQ; 2.
 CC Pfam: PF00063; myosin_head; 1.
 CC Pfam: PF02736; Myosin_N; 1.
 CC Pfam: PF01576; Myosin_tail; 1.
 CC PRINTS: PR00193; MYOSINHEAVY.
 CC ProDom: PD000355; myosin_head; 1.
 CC SMART: SM00015; IQ; 1.
 CC SMART: SM00242; MYSC; 1.
 CC PROSITE: PS50096; IQ; 1.
 CC
 CC Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family;
 KW Calmodulin-binding.
 FT DOMAIN 1 782 MYOSIN HEAD-LIKE.
 FT DOMAIN 783 812 IQ.

FT	DOMAIN	842	1939	COILED COIL (POTENTIAL).
FT	DOMAIN	1439	1443	POLY-ALA.
FT	NP_BIND	178	185	ATP (POTENTIAL).
FT	DOMAIN	659	679	ACTIN-BINDING.
FT	DOMAIN	759	773	ACTIN-BINDING.
FT	MOD_RES	129	129	METHYLATION (TRI-) (POTENTIAL).
FT	MOD_RES	697	697	ALKYLATION (SH-1) (POTENTIAL).
FT	MOD_RES	707	707	ALKYLATION (SH-2) (POTENTIAL).
FT	CONFLICT	1633	1633	Q -> L (IN REF. 2).
FT	CONFLICT	1651	1651	H -> Q (IN REF. 2).
FT	CONFLICT	1686	1687	EL -> DV (IN REF. 2).
FT	CONFLICT	1693	1693	V -> G (IN REF. 2).
FT	CONFLICT	1844	1844	K -> R (IN REF. 2).
FT	CONFLICT	1879	1879	A -> T (IN REF. 2).
FT	CONFLICT	1885	1885	E -> Q (IN REF. 2).
FT	CONFLICT	1907	1907	E -> V (IN REF. 2).
FT	CONFLICT	1928	1928	D -> N (IN REF. 2).
FT	CONFLICT	1933	1935	OKM -> KR (IN REF. 2).
SQ	SEQUENCE	1939	AA; 223626	MW; DBC8297/DFE83115A CRC64;

Query Match 4.4%; Score 171; DB 1; Length 1939;
 Best Local Similarity 23.2%; Pred. No. 0.2;
 Matches 89; Conservative 61; Mismatches 130; Indels 104; Gaps 17;

Qy	384	IKVSKMEQKFRMLSDAPTVKESCKTSSNNMSVNTLAKRIPNYQLS---PTKLPSPIN	440
Db	842	LKSAETEKEMANKEEGRVKESEKSEARRKELEKEMVSLQEKNDLQFQVQABQDNLN	901
Qy	441	KSQDRASQOQNTSIRNVFOPSTKKRE-----RDEE--NQEMSSCKSARIETSCS-LLEQ	492
Db	902	DAEERCQ-----LIKNIQLEAKVKEMTERLEDEEENNAELTS-KKKRKEDECESELKDD	955
Qy	493	TPATPSPSLWK-NKEQHLSENEPVDNTSDNNLFTDTDLKASIVKNSASKSHAAEKLRSNKK-	550
Db	956	IDDLLETLAKVEKEKHATENK-----VKNLTEEMAGLDEIIAKLTKEKKA	1000
Qy	551	-----REMDVATEDEVLEQLFDTKPELEIDVKKVQKEEDVNVKRPMDIETNDFTS	604
Db	1001	LQEAHQALDQLQAEEDKVNTL---TKSVKLEQVQDLEGLSEKQKVRMDLER-----	1052
Qy	605	DEAVPESSKISQENIGKREKELKEDSLWSAKEISN--NDKLQDDSEMLPKLLLTFRSL	662
Db	1053	-----AKRLEGDLNVQESIMDLNKLQ-----LEELKKKEF---1087	
Qy	663	VIKNSTSRNPSPGINDYQ-QLNFKFKKVTYPGAGKLPHTIGGSLIAHARKNTELEE	721
Db	1088	----DISOONSKIEDEQALALQLOKKE-----NOARTEEELEE	1122
Qy	722	WLQEMEYQNHAKESLADDLFR	745
Db	1123	ELEAE---RTARAKVEKLRSDLTR	1143

RESULT 5
 MYH6_HUMAN
 ID MYH6_HUMAN STANDARD: PRT; 1939 AA.
 AC P13533; Q13943; Q14906; Q14907;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain, cardiac muscle alpha isoform (MyHC-alpha).
 GN MYH6 OR MYHA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 RX MEDLINE=92133665; PubMed=1776652;
 RA Matsuo K., Beisel K.W., Furutani M., Arai S., Takao A.;
 "Complete sequence of human cardiac alpha-myosin heavy chain gene and
 amino acid comparison to other myosins based on structural and

functional differences.";
 RL Am. J. Med. Genet. 41:537-547(1991).
 [2]
 RP MEDLINE=94140346; PubMed=8307559;
 RX Epp T.A., Dixon I.M., Wang H.Y., Sole M.J., Liew C.-C.;
 RA "Structural organization of the human cardiac alpha-myosin heavy
 chain gene (MYH6).";
 RT Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).
 RL Genomics 18:505-509(1993).
 [3]
 RP MEDLINE=88299163; PubMed=2969919;
 RX Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;
 RA "Molecular cloning and characterization of human cardiac alpha- and
 beta-form myosin heavy chain complementary DNA clones. Regulation of
 expression during development and pressure overload in human
 atrium.";
 RT J. Clin. Invest. 82:524-531(1988).
 RL J. FUNCTION: MUSCLE CONTRACTION.
 CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
 MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; D00943; BAA00791.1; -;
 DR EMBL; Z20656; CAA79675.1; -;
 DR EMBL; M25140; AAA60386.1; -;
 DR EMBL; M25162; AAA60386.1; JOINED.
 DR EMBL; M25142; AAA60387.1; -;
 DR EMBL; M25141; AAA60387.1; JOINED.
 DR EMBL; M21664; AAA36344.1; -;
 DR PIR; A46762; A46762.
 DR PIR; A28908; A28908.
 DR RSP; P08799; 1MMD.
 DR MIM; 160710; -;
 DR InterPro; IPR000048; IQ.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00612; IQ; 2.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF02736; Myosin_N; 1.

DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS50096; IQ; 1.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family;
 KW Calmodulin-binding.
 FT DOMAIN 1 782 MYOSIN HEAD-LIKE.
 FT DOMAIN 783 812 IQ.
 FT DOMAIN 842 1939 COILED COIL (POTENTIAL).
 FT NP_BIND 178 185 ATP.
 FT DOMAIN 657 679 ACTIN-BINDING.
 FT DOMAIN 759 773 ACTIN-BINDING.
 FT MOD_RES 129 129 METHYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 697 697 ALKYLATION (SH-2) (POTENTIAL).
 FT CONFLICT 88 88 Q -> E (IN REF. 1).
 FT CONFLICT 574 574 Q -> P (IN REF. 1).
 FT CONFLICT 608 608 A -> G (IN REF. 1).
 FT CONFLICT 744 744 T -> A (IN REF. 1).
 FT CONFLICT 790 790 M -> I (IN REF. 1).
 FT CONFLICT 1014 1014 V -> A (IN REF. 1).
 FT CONFLICT 1021 1021 S -> T (IN REF. 1).
 FT CONFLICT 1101 1101 A -> V (IN REF. 1).
 FT CONFLICT 1290 1290 A -> S (IN REF. 1).
 FT CONFLICT 1373 1373 W -> C (IN REF. 1).
 FT CONFLICT 1533 1533 L -> N (IN REF. 5).
 FT CONFLICT 1540 1540 L -> M (IN REF. 5).
 FT CONFLICT 1577 1578 KL -> NV (IN REF. 5).
 FT CONFLICT 1705 1706 EQ -> DR (IN REF. 1).
 FT CONFLICT 1733 1733 E -> D (IN REF. 1).
 FT CONFLICT 1734 1734 A -> S (IN REF. 2).
 FT CONFLICT 1737 1737 T -> S (IN REF. 1).
 FT CONFLICT 1763 1763 D -> H (IN REF. 1).
 FT CONFLICT 1788 1788 M -> I (IN REF. 3).
 FT CONFLICT 1871 1871 R -> N (IN REF. 5).
 FT CONFLICT 1882 1882 D -> G (IN REF. 5).
 FT CONFLICT 1890 1890 Q -> R (IN REF. 5).
 FT CONFLICT 1933 1933 MISSING (IN REF. 5).
 SQ SEQUENCE 1939 AA; 223689 MW; ECB87E7CE8768B6F CRC64;
 Query Match 4.4%; Score 170.5; DB 1; Length 1939;
 Best Local Similarity 23.4%; Pred. No. 0.21;
 Matches 91; Conservative 57; Mismatches 132; Indels 109; Gaps 17;
 QY 388 KMEQKPMQLSQDAPTVKESCKTSSNNMVSNTLAKMRIPNYQLSPKLPISINKSDRAS 447
 Db KLYFKIKPLKLSAETEKENA-TMKEEFGRKETLEKSEARRKELE-EKMSVLIQEKNDLQ 890
 QY 448 QQQQTN-----IRNYFQPSYTKR---ERDEENQEMS---SCKSARIETSCS 488
 Db LQVQAEQDNLDAEERCDOGLINKIOLEAKVKEMNERLEDEEPMNAELTAKRKLEDECS 950
 QY 489 -LLEQTQPATPSLWK-NKEOHLSENEPVDNTSDNNLTDTDLKSIYKNSASKSHAAEKL 546
 Db ELKKDIDDLLETLAKVEKEKHATENK-----VKNLTTEEMAGLDEITAKLT 995
 QY 547 SNRK-----REMDVAIIDEVLEQLFKDTPKELEIDYKVKQEQEDVNVKRRPRMDIET 599
 Db KKKALQEAHQALDLDQVEEDKVNLSL---SKSVKLEQVDDLEGSLEQEKVKRMDLER 1052
 QY 600 NDTFDEAVPESSKISQENEIGKREKREKEDSLWSAKEISN---NDKLQDSEMLPKLLLT 657
 Db 1053 -----AKRKGDLKLTQESIMDLENDKLQ-----LEELKKKK 1085
 QY 658 EFRSLVKNSTSRNPSGINDDYG-OLKNFKKFKKVTYPGAGKLPHIIGSGDLIAHARKN 716
 Db 1086 EF-----DINQNSKIEDEQALALQLOKKLKE-----NQARI 1117
 QY 717 TELEEWLRQEMEVQNHAKESLADDLFR 745

CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
CC OF MAP1B.
CC
CC -1- SIMILARITY: TO MAP1A.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X51396; CAA35761.1; -.
DR PIR: S07549; QRMSP1.
DR MGD: MGI:1306778; Mtap1b.
DR InterPro: IPR000102; MAP1B_neuraxin.
DR Pfam: PF00414; MAP1B_neuraxin; 10.
DR PROSITE: PS00230; MAP1B_NEURAXIN; 7.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN ? 2464 MAP1 LIGHT CHAIN LC1.
FT REPEAT 1874 1890 MAP1B 1.
FT REPEAT 1891 1907 MAP1B 2.
FT REPEAT 1908 1924 MAP1B 3.
FT REPEAT 1925 1941 MAP1B 4.
FT REPEAT 1942 1958 MAP1B 5.
FT REPEAT 1959 1975 MAP1B 6.
FT REPEAT 1993 2009 MAP1B 7.
FT REPEAT 2010 2026 MAP1B 8.
FT REPEAT 2027 2043 MAP1B 9.
FT REPEAT 2044 2060 MAP1B 10.
FT DOMAIN 589 787
FT LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT KKEE AND KKEI/V REPEATS).
SQ SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFD8DA87 CRC64;

Query Match 4.4%; Score 170; DB 1; Length 2464;
Best local Similarity 18.08; Pred. No. 0.3;
Matches 171; Conservative 148; Mismatches 301; Indels 330; Gaps 44;
QY 6 PAAGGAGEPYRLITGVYVGRKNCALIEDQISIRNH---AVLTANFSV-----TNLS 58
DB 20 PAASTSLSHRFLDSKFY-----LLVVGVETVTEEHLRAIGNIELGIRSWDNL 71
QY 59 QTD---BIPVLTKDNKSKYGFVNEEKMQNGFSRTLK-----SGDITGVFGSKFRI 108
DB 72 ECNLDQELKLFVSRHSRSPVPGQKILHRSQVLETVVVLINPSDEAVS-----TEVRL 126
QY 109 -----EYEPVACSSCLDYSGKTALNQAILQLGGFTVNNWTECTHLVWVSVKVTI--- 159
DB 127 MITDAARKLLVLGTQCFTNG-----ELIQSGSFSFQFIETFDQIEIGELLSTTHPA 181
QY 160 -RTICALIC-----GRPIVKPEY-----FTBFLKAVQSKKQ 189
DB 182 NKASLTLCFCEGDKNKSNLDRHNLQDFINIKLNSASILPEMEGLSEFTEYLS--ESVEV 239
QY 190 PQIESYPP-----LDPE-----SIGSKNVDLSGRQKRIKFKGKTF 227
DB 240 PSPFDILEPPTSGGFLKSLKPCCVIFPGGRGDSALFAVNGFNLINGSERKSCF----- 294
QY 228 IFLNAKQHKLLSSAVFVGGE-----ARLITEENEEHNFFLAPGTCVVDITGN 277
DB 295 -WKLIHRLDRVDSLTHLTHIGDDNLPGINSMLQRIAELEER-----SQGSTS 341
QY 278 SQTLPDCQKWIQSDMLQRLRPIPEAIGLAVIFMTT-KNYCDPQGHSTGLKTT 336
DB 342 N-----SDWMKNLIS-----PD-----LGVVFLNVPENLKDPE----- 369
QY 337 TPGPSLSQGVSVDEKLMPSAPVNTTIVADTSEQADTWLSERPK----- 382
DB 370 ---PNIKMKRSIEACFTLOYLN-----KLSMKPEPLFRSVGNTIEPVI 410
QY 383 -----EIKVSKMEQKFRMLSDAPTAKESCKTSSNNNSV----- 417

DB 411 LFKQMGVGLKLEMYLVNPKVSSKEMQYF--MQQWTGNTKDKAEILILPNGQEVYDIPISYLS 468
QY 418 -----SNTLAKMRIPNYQLSPTKLPINSKSK-----DRASQOQNTSI--- 455
DB 469 VSSLIVVHPANPAEKIIRVLPFGNSTQYNILEGLEKLLKHLDFLQPLATKDLTGQVPTP 528
QY 456 -----RNYQPSTK-----KRERDEENQEMSSCKSARIETSCSLLLEQTQP 495
DB 529 PVKQVKLQRADSRSLKPKATPKVASKSVRKSKEETPEVT--KTSQVE----- 575
QY 496 ATPSLWNKKEOHLSENPEVDNSNNLFTDPLKISIVKNSASKSHAAEKLRSNKKREM-D 554
DB 576 KTKPVESKEVLVKDKAPVKTESKPSV-TEKEVSKESQSPVKAIEVAKQATESPKPKY 634
QY 555 DVAIEDEV---LEQ-----LFKDTKP-----ELEIDVKVO-KOEEDVNV 589
DB 635 DKVYKKEIKTKLEEKKEKPKKEVVKEDKTPLKDKDEPKREEVKKEIKKEIKKEEREL 694
QY 590 KRPMDIETNDTFSDEAVPESKISQENETGKREKEDSLWSAKEISNNKLDQ---D 646
DB 695 KKEVKKETPLKDA-KKEVKKEEKKEVKEEKEPKKEIKKIS---KDIKKSTPQSDTKKP 749
QY 647 SEMPLPKLLLTPEFRSLVINKSTSRNP--SGINDDYGQLKNEKPKKVTYPG-----AGK 698
DB 750 SALAPKVAKKEE-----STKKEPLAAGKLDKGKVKVYIKKEGKTTEAAATAVGTAA 801
QY 699 LPHIIGGSDLIHARKNTELEEWLRQEME-----VONQHAKEESLADDL 743
DB 802 TAAVVAAGTAASGPVKELEAERSILMSSPEDLTDKDFELKAEIDVAKDI 851
RESULT 8
BK1 YEAST
ID BK1 YEAST STANDARD; PRT; 1478 AA.
AC Q01389; P32894;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Serine/threonine protein kinase BK1/SLK1/SSP31 (EC 2.7.-.-).
GN BK1 OR SLK1 OR SSP31 OR LAS3 OR YJL095W OR J0906.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92186847; PubMed=1545797;
RA Costigan C., Gehring S., Snyder M.;
RT "A synthetic lethal screen identifies SLK1, a novel protein kinase
RT homolog implicated in yeast cell morphogenesis and cell growth.";
RL Mol. Cell. Biol. 12:1162-1178(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92104496; PubMed=1840547;
RA Irie K., Araki H., Oshima Y.;
RT "A new protein kinase, SSP31, modulating the SMP3 gene-product
RT involved in plasmid maintenance in Saccharomyces cerevisiae.";
RL Gene 108:139-144(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=EG123;
RX MEDLINE=92107166; PubMed=1729597;
RA Lee K.S., Levin D.E.;
RT "Dominant mutations in a gene encoding a putative protein kinase
RT (BCK1) bypass the requirement for a Saccharomyces cerevisiae protein
RT kinase C homolog.";
RL Mol. Cell. Biol. 12:172-182(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RX MEDLINE=95176706; PubMed=7871887;
RA Miosga T., Boles E., Schaaff-Gerstenschlaeger I., Schmitt S.,
RA Zimmermann F.K.;

—

FT	DOMAIN	1937	1989	LAMININ EGF-LIKE 19.	DISULFID	1937	1951	BY SIMILARITY.
FT	DOMAIN	1990	2036	LAMININ EGF-LIKE 20.	DISULFID	1939	1958	BY SIMILARITY.
FT	DOMAIN	2037	2083	LAMININ EGF-LIKE 21.	DISULFID	1961	1970	BY SIMILARITY.
FT	DOMAIN	2084	2131	LAMININ EGF-LIKE 22.	DISULFID	1973	1987	BY SIMILARITY.
FT	DOMAIN	2693	2884	LAMININ G-LIKE 1.	DISULFID	1990	2000	BY SIMILARITY.
FT	DOMAIN	2896	3066	LAMININ G-LIKE 2.	DISULFID	1992	2007	BY SIMILARITY.
FT	DOMAIN	3072	3235	LAMININ G-LIKE 3.	DISULFID	2009	2018	BY SIMILARITY.
FT	DOMAIN	3310	3482	LAMININ G-LIKE 4.	DISULFID	2021	2034	BY SIMILARITY.
FT	DOMAIN	3488	3669	LAMININ G-LIKE 5.	DISULFID	2037	2048	BY SIMILARITY.
FT	DISULFID	298	307	BY SIMILARITY.	DISULFID	2039	2055	BY SIMILARITY.
FT	DISULFID	300	320	BY SIMILARITY.	DISULFID	2057	2066	BY SIMILARITY.
FT	DISULFID	322	331	BY SIMILARITY.	DISULFID	2069	2081	BY SIMILARITY.
FT	DISULFID	334	354	BY SIMILARITY.	DISULFID	2084	2096	BY SIMILARITY.
FT	DISULFID	357	366	BY SIMILARITY.	DISULFID	2086	2103	BY SIMILARITY.
FT	DISULFID	359	391	BY SIMILARITY.	DISULFID	2105	2114	BY SIMILARITY.
FT	DISULFID	394	403	BY SIMILARITY.	DISULFID	2117	2129	BY SIMILARITY.
FT	DISULFID	406	424	BY SIMILARITY.	CARBOHYD	121	121	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	427	438	BY SIMILARITY.	CARBOHYD	140	140	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	429	445	BY SIMILARITY.	CARBOHYD	249	249	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	447	456	BY SIMILARITY.	CARBOHYD	351	351	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	459	469	BY SIMILARITY.	CARBOHYD	477	477	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	472	484	BY SIMILARITY.	CARBOHYD	511	511	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	474	491	BY SIMILARITY.	CARBOHYD	530	530	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	493	502	BY SIMILARITY.	CARBOHYD	634	634	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	505	516	BY SIMILARITY.	CARBOHYD	761	761	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	519	531	BY SIMILARITY.	CARBOHYD	1014	1014	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	521	538	BY SIMILARITY.	CARBOHYD	1341	1341	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	540	549	BY SIMILARITY.	CARBOHYD	1705	1705	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	552	561	BY SIMILARITY.	CARBOHYD	1756	1756	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	564	576	BY SIMILARITY.	CARBOHYD	1868	1868	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	566	583	BY SIMILARITY.	CARBOHYD	1944	1944	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	585	594	BY SIMILARITY.	CARBOHYD	1986	1986	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	597	607	BY SIMILARITY.	CARBOHYD	2002	2002	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	610	622	BY SIMILARITY.	CARBOHYD	2159	2159	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	612	629	BY SIMILARITY.	CARBOHYD	2207	2207	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	631	640	BY SIMILARITY.	CARBOHYD	2231	2231	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	643	653	BY SIMILARITY.	CARBOHYD	2235	2235	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	656	668	BY SIMILARITY.	CARBOHYD	2401	2401	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	658	674	BY SIMILARITY.	CARBOHYD	2421	2421	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	676	685	BY SIMILARITY.	CARBOHYD	2487	2487	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	688	698	BY SIMILARITY.	CARBOHYD	2821	2821	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	701	715	BY SIMILARITY.	CARBOHYD	3087	3087	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	703	724	BY SIMILARITY.	CARBOHYD	3242	3242	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	726	735	BY SIMILARITY.	CARBOHYD	3541	3541	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	DISULFID	738	753	BY SIMILARITY.	SEQUENCE	3672	3672	AA; 404223.MW; 28E262DB5FF14BFA CRC64;
FT	DISULFID	756	770	BY SIMILARITY.				
FT	DISULFID	758	777	BY SIMILARITY.				
FT	DISULFID	779	788	BY SIMILARITY.				
FT	DISULFID	791	806	BY SIMILARITY.				
FT	DISULFID	1415	1427	BY SIMILARITY.				
FT	DISULFID	1417	1434	BY SIMILARITY.				
FT	DISULFID	1436	1445	BY SIMILARITY.				
FT	DISULFID	1448	1458	BY SIMILARITY.				
FT	DISULFID	1461	1469	BY SIMILARITY.				
FT	DISULFID	1463	1476	BY SIMILARITY.				
FT	DISULFID	1478	1487	BY SIMILARITY.				
FT	DISULFID	1490	1503	BY SIMILARITY.				
FT	DISULFID	1506	1520	BY SIMILARITY.				
FT	DISULFID	1508	1527	BY SIMILARITY.				
FT	DISULFID	1529	1538	BY SIMILARITY.				
FT	DISULFID	1541	1551	BY SIMILARITY.				
FT	DISULFID	1554	1566	BY SIMILARITY.				
FT	DISULFID	1556	1573	BY SIMILARITY.				
FT	DISULFID	1575	1584	BY SIMILARITY.				
FT	DISULFID	1587	1602	BY SIMILARITY.				
FT	DISULFID	1830	1839	BY SIMILARITY.				
FT	DISULFID	1832	1846	BY SIMILARITY.				
FT	DISULFID	1849	1858	BY SIMILARITY.				
FT	DISULFID	1861	1877	BY SIMILARITY.				
FT	DISULFID	1880	1894	BY SIMILARITY.				
FT	DISULFID	1882	1905	BY SIMILARITY.				
FT	DISULFID	1907	1916	BY SIMILARITY.				
FT	DISULFID	1919	1934	BY SIMILARITY.				

Query Match 4.3%; Score 168.5; DB 1; Length 3672;
Best Local Similarity 20.4%; Pred. No. 0.6;
Matches 127; Conservative 105; Mismatches 201; Indels 189; Gaps 33;

QY	238	LSSAVFVGGEAR	-----LITEENEENHF-----	FLAPGTCVDTGI	275
DB	2160	ISSATIVGARLARNKKEFN	INEITKMLNDENSGNVFGDAQDILLNSTQIQKLVRTK	2219	
QY	276	TNSQTLIPDCQK	-----KWIQSIMDLQ--	ROGLRPIPEAEIGLAVIFMTTKNYCDPQG	327
DB	2220	THSONSVSSAKNITLNGTE	FLQEVVKRAQARSVSL--	AEIALAI-----	2264
QY	328	HPSTGLKTTTGPSPSGV	SQVDEKLMPSAPVNTT	Y---VADTESEQADT--	WDLSERPK 382
DB	2265	-----GSSKAVNVDPRL	KEABETLMTLEAASADQY	PERAKQTPVPGKLEETQK	2312
QY	383	EI-----KVKMEQKFR	-----MLSDAPT	VKESCKTSSNNNSVNTLAKMRI	426
DB	2313	KIQETEKLDKQKETFEA	QKRAEELAYLSAQQLKES-	KSADKSNNAKMLQJTKV	2371
QY	427	PN-----YQLSPTKLP	SINKS-KDRASO-----	QQOTNSIRNYF	459
DB	2372	ENLVAATDLDLREAR	AAKGBFQKLNVAIGNITEN	LKDKREMTHAVITLTNETRNDVAEL	2431
QY	460	QPSTKKRDEENQEMS-	--SCKSARIETSCSLLEQ	OTAPTPSLMKNKEQHLSENEPVDN	517

```
Db 2432 EAAKRVRRDEKSDVMQVLVNAKAHELHLQATTLRQT-----FDNN 2471
Qy 518 SDN-----NLF---TDT--DLKSIVKNSASKSHAAEKLRSNKKKREMDDDVAIEDEVLE 564
Db 2472 KDNTDQAVEAANAFSLNLTDLKNAKAQIDN-AYEALSAPAFASVQNAKDPPDETKE 2530
Qy 565 Q---LFKDTKPELEIDVKVQKEEDV-----NVRKR-----PRMDIETNDFSDAV 608
Db 2531 KIDALSKTVSODLKRETKLKKQLQELTSEKLRKRRAVAGIPKYSKNPLDST-DEKV 2589
Qy 609 PESSKISO-----NEI-GKKREL--REDSLSAKEISNNDKLODDSEMLPKK 653
Db 2590 QEVKLAEDIANTEETRAKISEIAGKAETKANSMEGIRLARRNSVO-LNK LAP-- 2646
Qy 654 LLLTFEFLSL-VIKNSTSRNPNGINDYQGLKFKFKKVTYPYGAGKLPFIHIGSDLIH 712
Db 2647 VIVSKFEELKLSARSAKVDSDKYSQI---KEMIAVARDAANRIK-----LGAH 2695
Qy 713 ARKNTLEELWLRQEMEVONQHA 734
Db 2696 FEGSSLDLNPQRVTRSAHA 2717

RESULT 11
TPR_HUMAN
ID TPR_HUMAN STANDARD; PRT; 2349 AA.
AC P12270;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nucleoprotein TPR.
GN TPR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=93064711; PubMed=1437155;
RX Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
RT "tpr, a large coiled coil protein whose amino terminus is involved in
RT activation of oncogenic kinases, is localized to the cytoplasmic
RT surface of the nuclear pore complex."
RL Oncogene 7:2329-2333(1992).
RN [2]
RP REVISTONS, AND CHARACTERIZATION.
RX MEDLINE=95096166; PubMed=7798308;
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
RT "tpr, a large coiled coil protein whose amino terminus is involved in
RT activation of oncogenic kinases, is localized to the cytoplasmic
RT surface of the nuclear pore complex."
RL J. Cell Biol. 127:1515-1526(1994).
RN [3]
RP SEQUENCE OF 1-142 FROM N.A.
RX MEDLINE=98262257; PubMed=3387099;
RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
RA "tpr homologues activate met and raf."
RL Oncogene 2:617-619(1988).
RN [4]
RP FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE
RN COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS
RN IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.
RN [5]
RP SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
RN COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
RN TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
RN COMPONENTS, INCLUDING P62.
RN [6]
RP TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND
RN BRAIN, LOWER LEVELS IN HEART, LIVER, AND KIDNEY.
RN [7]
RP DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK
RN OR RAF GENES.
RN [8]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X66397; CA47021.1; -.
DR EMBL; Y00672; CA68681.1; -.
DR PIR; S00928; S00928.
DR MIM; 189940; -.
KW Coiled coil; Proto-oncogene; Chromosomal translocation;
KW Nuclear protein; Transport.
FT DOMAIN 78 360 COILED COIL (POTENTIAL).
FT DOMAIN 422 571 COILED COIL (POTENTIAL).
FT DOMAIN 575 628 COILED COIL (POTENTIAL).
FT DOMAIN 758 805 COILED COIL (POTENTIAL).
FT DOMAIN 834 869 COILED COIL (POTENTIAL).
FT DOMAIN 934 979 COILED COIL (POTENTIAL).
FT DOMAIN 1004 1064 COILED COIL (POTENTIAL).
FT DOMAIN 1138 1166 COILED COIL (POTENTIAL).
FT DOMAIN 1196 1241 COILED COIL (POTENTIAL).
FT DOMAIN 1262 1304 COILED COIL (POTENTIAL).
FT DOMAIN 1354 1434 COILED COIL (POTENTIAL).
FT DOMAIN 1476 1595 COILED COIL (POTENTIAL).
FT DOMAIN 527 530 POLY-SER.
FT DOMAIN 1833 1836 POLY-GLU.
FT DOMAIN 1957 1964 POLY-ASP.
FT DOMAIN 2295 2298 POLY-SER.
SQ SEQUENCE 2349 AA; 265600 MW; AFDD6885CEDCA9EF CRC64;

Query Match 4.3%; Score 167.5; DB 1; Length 2349;
Best Local Similarity 22.1%; Pred. No. 0.39;
Matches 96; Conservative 85; Mismatches 183; Indels 71; Gaps 21;

Qy 333 LKTTTPGP---SLS-QGVSVDEKLMPSAVNTTYYVADTESEQADTWLSEPRKEIKVSK 388
Db 1174 VKGVQGLNVLSEBEGKSQEQILEILRFIRREKEIAETFEVAQVESLRVQ---RVEL 1230
Qy 389 MEQFRMLSDAPTVKESCKTSSNNNSMVSNTLAKMRIPNYQLSPTKLPISNKSDRASQ 448
Db 1231 LERELQELSDLSNAERKQVQVTAQMAQHEELMKKTETMNVVMTNKM--LREKERLEQ 1288
Qy 449 QQQNSTNRYQPTSTKKRERD-----EENQEMSSCKSARIETSCSLLEQTOPATPSLWKN 503
Db 1289 DLQ-----QMCAKVRKLELDILPLQEAANELSE-KSGMLQAEKKLLEE---DVKRKA 1337
Qy 504 KEQHL-SENEPVDTSNNDNLFTDLDL--KSI-----VKNSASKSHAA-----EKLRS 547
Db 1338 RNQHLVSOQKDPDTEEVYRKLLSEKEVHTKRIQQLTEELGRKAEIARSNASLTNNQNLIQ 1397
Qy 548 NKKREMDDDVAIEDEVLEQLFKDTKPE-LEIDVKVQKEEDVNVKRRPRMDIETNDFSD 606
Db 1398 SLKEDLNKVRTEKETIQ---KDLDAKIIDIQEKVKTITQVKKIGRRYKTYQVEELKAQDK 1454
Qy 607 AVPESSKISQENE-----IGKKRELKEDSLWSAKEISNNDKLODDSEMLPKK--LLTEF 659
Db 1455 VMETSQSSGDHQHVSQVQEMQELKE-TLNQAE--TKSKLSLESQVENLQTLSEKETE 1511
Qy 660 RSL-----VIKNSTSRNPNGINDYQGLKFKFKKVTYPYGAGKLPFIHIGSDLIH 706
Db 1512 RNQEQVTQVQSELSRLNQLDQRTTQEEQLRQKITEKEKTRKAIWAASAKIAHLGK 1571
Qy 707 DLIAHARKNTELEE 721
Db 1572 DQL---TKENEELAQ 1583

RESULT 12
ID ATRX_CAEEL STANDARD; PRT; 1359 AA.
AC Q9U7E0; O02061;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
```


RN [7] SEQUENCE OF 1410-1935 FROM N.A.
 RP MEDLINE=88299163; PubMed=2969919;
 RA Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;
 RT "Molecular cloning and characterization of human cardiac alpha- and
 RT beta-form myosin heavy chain complementary DNA clones. Regulation of
 RT expression during development and pressure overload in human
 RT atrium.";
 RL J. Clin. Invest. 82:524-531(1988).
 RN [8]
 RN SEQUENCE OF 785-1935 FROM N.A.
 RP TISSUE-Skeletal muscle;
 RC MEDLINE=90235862; PubMed=1601980;
 RX Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
 RA Arnold H.H.;
 RT "Identification of three developmentally controlled isoforms of human
 RT myosin heavy chains.";
 RL Eur. J. Biochem. 189:55-65(1990).
 RN [9]
 RN SEQUENCE OF 1393-1935 FROM N.A.
 RP MEDLINE=87192738; PubMed=3032769;
 RX Jandreski M.A., Liew C.-C.;
 RA "Construction of a human ventricular cDNA library and
 RT characterization of a beta myosin heavy chain cDNA clone.";
 RL Hum. Genet. 76:47-53(1987).
 RN [10]
 RN REVIEW ON VARIANTS.
 RP MEDLINE=96039076; PubMed=8533830;
 RX Arai S., Matsuoka R., Hirayama K., Sukurai H., Tamura M., Ozawa T.,
 RA Kimura M., Imanura S.-I., Furutani Y., Joh-O K., Kawana M., Takao A.,
 RA Hosoda S., Momma K.;
 RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in
 RT hypertrophic cardiomyopathy.";
 RL Am. J. Med. Genet. 58:267-276(1995).
 RN [11]
 RN VARIANTS CMH1 GLU-256 AND ARG-741.
 RP MEDLINE=93248216; PubMed=8483915;
 RX Fananapazir L., Dalakas M.C., Cyran F., Cohn G., Epstein N.D.;
 RA "Missense mutations in the beta-myosin heavy-chain gene cause central
 RT core disease in hypertrophic cardiomyopathy.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:3993-3997(1993).
 RN [12]
 RN VARIANT CMH1 GLN-403.
 RP MEDLINE=90367131; PubMed=1975517;
 RX Geisterfer-Lowrance A.A.T., Kass S., Tanigawa G., Vosberg H.-P.,
 RA McKenna W., Seidman C.E., Seidman J.G.;
 RA "A molecular basis for familial hypertrophic cardiomyopathy: a beta
 RT cardiac myosin heavy chain gene missense mutation.";
 RL Cell 62:999-1006(1990).
 RN [13]
 RN VARIANTS CMH1 GLN-249; GLN-403; CYS-453 AND MET-606.
 RP MEDLINE=92204193; PubMed=1552912;
 RX Watkins H., Rosenzweig A., Hwang D.S., Levi T., McKenna W.,
 RA Seidmann C.E., Seidmann J.G.;
 RT "Characteristics and prognostic implications of myosin missense
 RT mutations in familial hypertrophic cardiomyopathy.";
 RL New Engl. J. Med. 326:1108-1114(1992).
 RN [14]
 RN VARIANTS CMH1 GLN-403; CYS-453; ARG-584 AND MET-606.
 RP MEDLINE=94070863; PubMed=8250038;
 RX Watkins H., Thierfelder L., Anan R., Jarcho J., Matsumori A.,
 RA McKenna W., Seidman J.G., Seidman C.E.;
 RT "Independent origin of identical beta cardiac myosin heavy-chain
 RT mutations in hypertrophic cardiomyopathy.";
 RL Am. J. Hum. Genet. 53:1180-1185(1993).
 RN [15]
 RN VARIANTS CMH1 GLN-403 AND VAL-908.
 RP MEDLINE=92346810; PubMed=1638703;
 RX Epstein N.D., Cohn G.M., Cyran F., Fananapazir L.;
 RA "Differences in clinical expression of hypertrophic cardiomyopathy
 RT associated with two distinct mutations in the beta-myosin heavy chain
 RT gene. A 908Leu-->Val mutation and a 403Arg-->Gln mutation.";
 RL Circulation 86:345-352(1992).

RN [16]
 RN VARIANTS CMH1 LEU-403 AND TRP-403.
 RP MEDLINE=94075629; PubMed=8254035;
 RX Dausse E., Komajda M., Feller L., Dubourg O., Dufour C., Carrier L.,
 RA Wisniewsky C., Bercevic J., Hengstenberg C., Al-Mahdawi S.;
 RT "Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and
 RT identification of a hot spot for mutations in the beta-myosin heavy
 RT chain gene.";
 RL J. Clin. Invest. 92:2807-2813(1993).
 RN [17]
 RN VARIANT CMH1 TRP-403.
 RP MEDLINE=94093568; PubMed=8268932;
 RX Moolman J.C., Brink P.A., Corfield V.A.;
 RA "Identification of a new missense mutation at Arg403, a Cpg mutation
 RT hotspot, in exon 13 of the beta-myosin heavy chain gene in
 RT hypertrophic cardiomyopathy.";
 RL Hum. Mol. Genet. 2:1731-1732(1993).
 RN [18]
 RN VARIANT CMH1 ASN-615.
 RP MEDLINE=93038688; PubMed=1417858;
 RX Nishi H., Kimura A., Harada H., Toshima H., Sasazuki T.;
 RA "Novel missense mutation in cardiac beta-myosin heavy chain gene
 RT found in a Japanese patient with hypertrophic cardiomyopathy.";
 RL Biochem. Biophys. Res. Commun. 188:379-387(1992).
 RN [19]
 RN VARIANT CMH1 GLY-778.
 RP MEDLINE=93343938; PubMed=8343162;
 RX Harada H., Kimura A., Nishi H., Sasazuki T., Toshima H.;
 RA "A missense mutation of cardiac beta-myosin heavy chain gene linked
 RT to familial hypertrophic cardiomyopathy in affected Japanese
 RT families.";
 RL Biochem. Biophys. Res. Commun. 194:791-798(1993).
 RN [20]
 RN VARIANT CMH1 VAL-908.
 RP MEDLINE=93168485; PubMed=8435239;
 RX Al-Mahdawi S., Chamberlain S., Cleland J., Nihoyannopoulos P.,
 RA Gilligan D., French J., Choudhury L., Williamson R., Oakley C.;
 RT "Identification of a mutation in the beta cardiac myosin heavy chain
 RT gene in a family with hypertrophic cardiomyopathy.";
 RL Br. Heart J. 69:136-141(1993).
 RN [21]
 RN VARIANT CMH1 TRP-719.
 RP MEDLINE=95179132; PubMed=7874131;
 RX Greve G., Bachinski L., Friedman D.L., Czernuzewicz G., Anan R.,
 RA Towbin J., Seidman C.E., Roberts R.;
 RT "Isolation of a de novo mutant myocardial beta MHC protein in a
 RT pedigree with hypertrophic cardiomyopathy.";
 RL Hum. Mol. Genet. 3:2073-2075(1994).
 RN [22]
 RN VARIANTS CMH1 CYS-513; ARG-716 AND TRP-719.
 RP MEDLINE=94110336; PubMed=8282798;
 RX Anan R., Greve G., Thierfelder L., Watkins H., McKenna W., Solomon S.,
 RA Vecchio C., Shono H., Nakao S., Tanaka H., Mares A. Jr., Towbin J.A.,
 RA Spirito P., Roberts R., Seidman J.G., Seidman C.E.;
 RT "Prognostic implications of novel beta cardiac myosin heavy chain gene
 RT mutations that cause familial hypertrophic cardiomyopathy.";
 RL J. Clin. Invest. 93:280-285(1994).
 RN [23]
 RN VARIANT CMH1 THR-797.
 RP MEDLINE=96047159; PubMed=7581410;
 RX Moolman J.C., Brink P.A., Corfield V.A.;
 RA "Identification of a novel Ala797Thr mutation in exon 21 of the beta-
 RT myosin heavy chain gene in hypertrophic cardiomyopathy.";
 RL Hum. Mutat. 6:197-198(1995).
 RN [24]
 RN VARIANT CMH1 CYS-453.
 RP MEDLINE=96209901; PubMed=8655135;
 RX Ko Y.-L., Chen J.-J., Tang T.-K., Cheng J.-J., Lin S.-Y., Liou Y.-C.,
 RA Kuan P., Wu C.-W., Lien W.-P., Liew C.-C.;
 RT "Malignant familial hypertrophic cardiomyopathy in a family with a
 RT 453Arg-->Cys mutation in the beta-myosin heavy chain gene;
 RT coexistence of sudden death and end-stage heart failure.";
 RL Hum. Genet. 97:585-590(1996).

RESULT 14

```

QY 369 SEQADTWDLSEPKKIKVSKMEOKFRMLSDQADPTVKESCTSSNNNSWVNTLAK----- 423
Db 195 SKDSISVSIKHOKKRYKSDSKE---MDFESTSTKRYKSSKNNSDAVSETGCKSNSSNV 251
QY 424 -----MRIPNVLQS-----PTKLPSINKSKDRASQOQOQNSIRNYFQPS 462
Db 252 GLMVHGLGSDAESWEPDMWLKNTSQNNSKKPTTKDAKKDKGSGSDAESVDS----- 303
QY 463 TKRREDEENQESSKARIETSCSLE-----QTOPATPFLMKNKEQHLSENPEPVDTN 517
Db 304 -KDAKKDKGKATDKTKGAKKDTTESTDAESGDSKDAKKGKESKDKKKDAKDAASDAE 362
QY 518 SDNNLFTDTOLKSIVKNASAKSHAAEKLNRKNKREMDVDAIEVLEQLFKDKYKPELEID 577
Db 363 SG-----DSKDAKKD--SKGKDKSKDKNKKDAKKDAKDAESTDAESGDSKDAKKSKG 413
QY 578 VKYQOEEDVNVKRRPRMDJETNDTFSDEA--VPESKISQENEICKKRELKEDSL-WSA 634
Db 414 KDSKDKDK---KDAKKDAESTDAESGDSKNAKKDKSKGKKDD--KKDAKKDAVSTDA 468
QY 635 KEISNND--KLQDDSEMLPKKKLLITFRSLVINKSTSRNPNGINDDYGQLKNEFKFKKVT 692
Db 469 DSESEGDAKSKDKSKDKKDLKKDKQKPKAMK---SKESTETESDWESKKVKRDSKDDT 525
QY 693 YPGAGKLPHIIGG-SOLIAHARKNTELEWLRQEMEVQNHAKESLADDLFRYPNLYK 751
Db 526 KKTAKKATSSGAESDVSSRYLKKT-----EMFKSDAEESESL-----FKPGSK 571
QY 752 RR 753
Db 572 KR 573

RESULT 15
MYH6 RAT
ID MYH6_RAT STANDARD; PRT; 1938 AA.
AC P02563; Q63351;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle alpha isoform (MyHC-alpha).
GN MYH6.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=90016822; PubMed=2798111;
RA Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
RT "Complete nucleotide sequence of full length cDNA for rat alpha
RT cardiac myosin heavy chain."
RL Nucleic Acids Res. 17:7527-7528(1989).
RN [2]
DISCUSSION OF SEQUENCE.
RX MEDLINE=90133919; PubMed=2614840;
RA McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
RT "Full-length rat alpha and beta cardiac myosin heavy chain sequences.
RT Comparisons suggest a molecular basis for functional differences."
RL J. Mol. Biol. 210:665-671(1989).
RN [3]
SEQUENCE OF 1-167 FROM N.A.
RX MEDLINE=84194059; PubMed=6585819;
RA Mahdavi V., Chambers A.P., Nadal-Ginard B.;
RT "Cardiac alpha- and beta-myosin heavy chain genes are organized in
RT tandem."
RL Proc. Natl. Acad. Sci. U.S.A. 81:2626-2630(1984).
RN [4]
SEQUENCE OF 1512-1938 FROM N.A.
RX MEDLINE=82220036; PubMed=7045682;
RA Mahdavi V., Periasamy M., Nadal-Ginard B.;
RT "Molecular characterization of two myosin heavy chain genes expressed

```

```

RT in the adult heart.";
RL Nature 297:659-664(1982).
RN [5]
SEQUENCE OF 1872-1938 FROM N.A.
RC STRAIN=Wistar; TISSUE=Heart;
RX MEDLINE=85179510; PubMed=6241892;
RA Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;
RT "Cardiac myosin heavy chain isozymic transitions during development
RT and under pathological conditions are regulated at the level of mRNA
RL Eur. Heart J. 5:181-191(1984).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE. SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X15938; CAA34064.1; -
CC EMBL: K01464; AAA41648.1; -
CC EMBL: J00751; AAA41653.1; -
CC EMBL: M32697; AAA41658.1; -
CC EMBL: S06005; S06005.
CC PIR: A02988; A02988.
CC HSSP: P08799; LMMD.
CC InterPro: IPR000048; IQ.
CC InterPro: IPR004009; Myosin_N.
CC InterPro: IPR002928; Myosin_tail.
CC InterPro: IPR001609; myosin_head.
CC Pfam: PF00612; IQ; 2.
CC Pfam: PF00063; myosin_head; 1.
CC Pfam: PF02736; Myosin_N; 1.
CC Pfam: PF01576; Myosin_tail; 1.
CC PRINTS: PR00193; MYOSINHEAVY.
CC ProDom: PD000355; myosin_head; 1.
CC SMART: SM00015; IQ; 1.
CC SMART: SM00242; MYSC; 1.
CC PROSITE: PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family;
KW Calmodulin-binding.
FT DOMAIN 1 781 MYOSIN HEAD-LIKE.
FT DOMAIN 782 811 IQ.
FT DOMAIN 842 1938 COILED COIL (POTENTIAL).
FT NP_BIND 177 184 ATP.
FT DOMAIN 656 678 ACTIN-BINDING.
FT DOMAIN 758 772 ACTIN-BINDING.
FT DOMAIN 789 806 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 815 832 CALMODULIN-BINDING (BY SIMILARITY).
FT MOD_RES 128 128 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 696 696 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 705 705 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 13 13 R -> AP (IN REF. 3).

```

```
FT CONFLICT 46 46 V -> A (IN REF. 3).
FT CONFLICT 51 52 VS -> AP (IN REF. 3).
FT CONFLICT 87 87 E -> Q (IN REF. 3).
FT CONFLICT 109 109 MISSING (IN REF. 3).
FT CONFLICT 1566 1566 F -> FF (IN REF. 4).
FT CONFLICT 1575 1575 R -> S (IN REF. 4).
FT CONFLICT 1721 1721 N -> T (IN REF. 4).
FT CONFLICT 1852 1852 T -> N (IN REF. 4).
FT CONFLICT 1870 1870 D -> N (IN REF. 4).
FT CONFLICT 1934 1934 M -> I (IN REF. 4 AND 5).
SQ SEQUENCE 1938 AA; 223507 MW; D7BD33FC2B19E3C2 CRC64;

Query Match 4.2%; Score 165; DB 1; Length 1938;
Best Local Similarity 22.5%; Pred. No. 0.41;
Matches 91; Conservative 61; Mismatches 123; Indels 130; Gaps 17;

QY 365 ADTSEQADTWLSERPK-----EIKVSKMEOKFRMLSQDAPTVKESCKTSSNNNSMWS 418
   :||| | : | | | : :||| | : : : : |
Db 844 AETEKEMANNKEFGRVKDALEKSEARKELEEKVSLQEKNDLQVQAEQDN----- 898
   : : : | | : : : | | | : | | : |||

QY 419 NTLAKMRIPNYQLSPTKLPKINSKDRASQQOQTNSIRNYFQPSKKR---ERDEENQEM 475
   : : : | | : : : | | | | | : : |||
Db 899 -----LADAEERCDO-----LIKNIQLEAKVKEMTERLEDEEEM 933

QY 476 S----SCKSARIETCS-LLEQTQTPATPSLWK-NKEQHLSENEPPVDNNDNLFDTDLKS 530
   : | : | | | : | | : | | : ||| : | :
Db 934 NAELTAKKRKLEDECSSELKDDIDLELTAKVEKEKHATENK-----VKN 978

QY 531 IVKNSASKSHAAEKLRSNKK-----REMDVAIEDVLEQLFKDTKPELEIDVKVQKQ 583
   : : | | | | | : : ||| : : : : |
Db 979 LTEEMAGLDIEIIAKLTKEKKALQEAHQALDDLQAEEDKVNTL---TKSKVKLEQQVDDL 1035

QY 584 EEDVNVKRPRMDIETNDTFSDEAVPESSKISQENETGKKRELKEDSLWSAKEISN--ND 641
   | : : | | | : | : ||| : | : : | |
Db 1036 EGSLEQEKVPRMDLER-----AKRLEGDKLTQESIMDLEND 1073

QY 642 KLQDDSEMLPKLLLTFRSLVIKNSTSRNPNGINDYG-OLKNFKFKRVKVTYPGAGKLP 700
   ||| | : | | | : | | : | : | : | :
Db 1074 KLQ-----LEEKLKKEF-----DISQNSKIEDEQALALQOKKKE----- 1111

QY 701 HIIGSDLIAHARKNTELEEWLRQEMEVQNHAKESLADDLFR 745
   : : |||| | : | | | | |
Db 1112 -----NOARIEEELEEELEAE---RTARAKVEKLRSDLTR 1142
```

Search completed: August 15, 2002, 01:49:33
Job time: 488 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:31:20 ; Search time 84.59 seconds
(without alignments)
12785.487 Million cell updates/sec

Title: US-09-837-602-1

Perfect score: 4403

Sequence: 1 ttccggcacaggcgcgttg.....accgggtggagctccagct 4403

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	79.8	1.8	7218	1	US-08-232-463-14
C 2	65	1.5	19124	2	US-08-487-826B-13
C 3	59	1.3	7218	1	US-08-232-463-14
C 4	59	1.3	19124	2	US-08-487-826B-13
C 5	57.2	1.3	665	2	US-08-883-795A-36
C 6	55.6	1.3	3023	4	US-09-308-022-4
C 7	54.8	1.2	406	3	US-08-928-799A-1
C 8	54.8	1.2	545	4	US-09-227-357-125
C 9	54.8	1.2	752	4	US-08-376-259-108
C 10	54.8	1.2	849	3	US-08-646-538-14
C 11	54.8	1.2	849	4	US-09-503-222-14
C 12	54.8	1.2	912	2	US-09-090-567-3
C 13	54.8	1.2	1378	1	US-08-075-533-20
C 14	54.8	1.2	1378	2	US-08-948-176-20
C 15	54.8	1.2	1378	5	PCT-US91-09160-20
C 16	54.8	1.2	1556	4	US-09-043-937A-3
C 17	54.8	1.2	1568	4	US-09-043-937A-1
C 18	54.8	1.2	1830	1	US-08-343-733A-2
C 19	54.8	1.2	1958	4	US-09-215-221-9
C 20	54.8	1.2	2082	4	US-09-293-505-9
C 21	54.8	1.2	2605	2	US-08-680-395-4
C 22	54.8	1.2	2775	1	US-08-730-771-1
C 23	54.8	1.2	2775	4	US-09-060-208-1
C 24	54.8	1.2	2961	4	US-08-446-935-6
C 25	54.8	1.2	3089	1	US-08-472-934-5
C 26	54.8	1.2	3089	2	US-08-323-460A-5
C 27	54.8	1.2	3089	2	US-08-461-146C-5

C 28	54.8	1.2	3089	3	US-08-461-145C-5
C 29	54.8	1.2	3089	4	US-08-628-829-9
C 30	54.8	1.2	3328	4	US-08-960-048-1
C 31	54.8	1.2	3699	3	US-08-646-538-6
C 32	54.8	1.2	3699	4	US-09-503-222-6
C 33	54.8	1.2	3792	2	US-08-992-334-1
C 34	54.8	1.2	3792	3	US-08-302-752-1
C 35	54.8	1.2	5234	2	US-08-992-334-2
C 36	54.8	1.2	5234	3	US-08-302-752-2
C 37	54.8	1.2	6045	3	US-08-675-566-18
C 38	54.8	1.2	6244	3	US-08-675-566-17
C 39	54.8	1.2	6447	3	US-08-675-566-16
C 40	54.8	1.2	6578	3	US-08-675-566-4
C 41	54.8	1.2	6612	3	US-08-675-566-15
C 42	54.8	1.2	6722	2	US-08-992-334-3
C 43	54.8	1.2	6722	3	US-08-302-752-3
C 44	54.8	1.2	6958	3	US-08-675-566-2
C 45	54.8	1.2	6994	3	US-08-675-566-1

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

[illegible]

RESULT 2

US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellem, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs

[illegible]

```

RESULT 3
US-08-232-463-14
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Davidson Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)883-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
CLONE: ptzgpt-Fls
US-08-232-463-14

Query Match 1.3%; Score 59; DB 1; Length 7218;
Best Local Similarity 10.1%; Pred. No. 0.00022;
Matches 45; Conservative 213; Mismatches 189; Indels 0; Gaps 0;

QY 3819 tctgtacacctctgtctttacaaaatttggtgtgtccagttattttcccttttaac 3878
Db 1083 YYYVYYY 1142
QY 3879 cmttcccaattcgggtgtgtagtggtgtccatttggttttaattgttatccct 3938
Db 1143 YYYVYYY 1202
QY 3939 gatagctataattgggtcagataaattctttacattcttagatgaagtccttgcgg 3998
Db 1203 YYYVYYY 1262
QY 3999 atacgtattgagataattacacctagctgtgtgcttgactgtttctttatgtctttg 4058
Db 1263 YYYVYYY 1322
QY 4059 atgaatagaagttttaaatttgcacaggtcacaattttttttctttgttgatatt 4118
Db 1323 YYYVYYY 1382
QY 4119 tttctctcaatttaaccaccaagatttcagatattctgctctattatataaaactttata 4178
Db 1383 YYYVYYY 1442
QY 4179 tttttattgttgatctacacctgaattgatgtatgttgtgaattatgatcagggtt 4238
Db 1443 TTCTTCTATCTCTTAACTACTTGCATAGATAGTAATTAACAGTGATGCCGTACATGCCGT 1502
QY 4239 cttttttcccccatacaagtatccag 4265
Db 1503 TTTTGGAACTGAATAGATCGCTCTAG 1529

RESULT 4
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5933827
; GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellem, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 1.3%; Score 59; DB 2; Length 19124;
Best Local Similarity 44.3%; Pred. No. 0.00031;
Matches 239; Conservative 0; Mismatches 300; Indels 0; Gaps 0;

QY 1618 attcagatcacagatttaaaattctattgtgaaaaattctgccagtaaatctcatgctgc 1677
Db 15449 AATGAAAAAAGAAAAATGAATATAAAAAAATTTTATATAAAAAAATTTTATATAAAAAA 15508
QY 1678 agaaagctaaagtaataaaaaaagggaatggatggtggtccatagaagatgaagt 1737
Db 15509 AAAAAAGGAGAAAAATTTTATAAAAAATATAAAAAATTAATAAAAAATATAAAAAATTTTGAT 15568
QY 1738 attggacagattattcaaggacacaaacccagattgataaattgattgaaagttcaaaa 1797
Db 15569 AGAATAAAAAATGAAGAAGATTATCAAAAAAATTAATAAAAAATTTTATATAAAAAA 15628
QY 1798 acaggaggagatgtcaattgtagaaaaagcccaagatggatagataagaacaatgacac 1857
Db 15629 AAATGATTATAAAAAAATAAAAAACAAAGAAAAAATTAATAAAAAAATTTTATATAAAAAA 15688
QY 1858 ttccagtgatgaagcagtagccagaagaatagcaaaatattctcaagaaatgaaattggaaa 1917
Db 15689 ATATATATATCAAAAAACAAAAAATTAATAAAAAATTTTATATAAAAAATTTATATCA 15748
QY 1918 gaaacgtgaactcaaggaagactcactatggtcagctcaagaaatattctcaacaatgacaa 1977
Db 15749 TAAATAAAAAAATTAATAAAAAATTTTATAAAAAAATTAATAAAAAATTTATATAAAAAA 15808
QY 1978 acttcaggatgatgtgagatgcttccaaaaaagctgttattgactgaattgactcact 2037


```

RESULT          9
US-08-976-259-108
; Sequence 108, Application US/08976259
; Patent No. 6316609
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Choi, Gil H.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: Nucleotide sequence of Escherichia coli
; Patent No. 6316609
; NUMBER OF SEQUENCES: 142
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2

```

APPLICANT: VONHARMS, JOHN N.
 TITLE OF INVENTION: Mutant Aequorea victoria Fluorescent
 TITLE OF INVENTION: Proteins Having Increased Cellular Fluorescence
 NUMBER OF INVENTIONS: 37
 NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, 8th Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/646,538
 FILING DATE: No. 6027881 yet assigned
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Weber, Kenneth A.
 REGISTRATION NUMBER: 31,677
 REFERENCE/DOCKET NUMBER: 015280-249000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 849 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

Query Match	1.2%	Score 54.8;	DB 4;	Length 849;
Best Local Similarity	83.3%;	Pred. No. 0.0009;		
Matches 50; Conservative	9;	Mismatches 1;	Indels 0;	Gaps 0;

